Sections 4.1 & 4.2 of the present work are slight modifications of content appearing in the following publication:

Adrian E. Radillo, Alan Veliz-Cuba, Zachary P. Kilpatrick, and Krešimir Josić. Evidence accumulation and change rate inference in dynamic environments. Neural Computation, 29(6):1561–1610, 2017.

We added the appropriate copyright notice at the start each concerned section and subsection.

OPTIMAL DECISION MAKING MODELS IN CHANGING ENVIRONMENTS

A Dissertation Presented to

the Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Adrian Ernesto Radillo

August 2018

OPTIMAL DECISION MAKING MODELS IN CHANGING ENVIRONMENTS

Adrian Ernesto Radillo

APPROVED:

Dr. Krešimir Josić, Chairman

Dr. Zachary P. Kilpatrick
University of Colorado Boulder

Dr. William Ott

Dr. Andrew Török

Dean, College of Natural Sciences and Mathematics

Acknowledgements

Special thanks ought to be given to the people who have been with me, supported me, and had an unbreakable faith in me, all along the journey of my Ph.D. I am thinking of both work and family relationships.

My adviser Krešimir Josić is the best I could have ever had. Not only has he closely followed my work from the beginning, but he also out-did himself in all the difficult times to make sure that I could finish the present work; he always had an ear to lend.

Professor Kilpatrick was also outstanding as a co-adviser. His constant availability and willingness to listen made him a precious ally.

My dear wife, parents, and brother, are, and always will be, a source of joy, strength, and happiness. And my dad nurtured the scientist in me.

Aside from the constant help that I received, there are also the sporadic ones, without which this work would lack much of its quality. To the people that gave it to me, I am extremely grateful.

Some of these people are fellow students: Simon, Robert, Dylan, Mo, Nick, Kazem, James, Kate, Alan, and Selahittin, from UH, Alex Piet from Princeton, and Brandon Williams from Rice. And some others are researchers, such as William Ott, Vaughn Climenhaga, and Bhargav Karamched from UH, and Alan Veliz-Cuba, Xaq Pitkow, and Joshua Gold, from other labs.

Finally, I would like to thank Linda Robinson, Avani Dave, and the rest of the staff from the department of Mathematics at UH for allowing me, through their

constant work, to arrive at this finalizing point in my studies.

OPTIMAL DECISION MAKING MODELS IN CHANGING ENVIRONMENTS

An Abstract of a Dissertation

Presented to

the Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Adrian Ernesto Radillo

August 2018

Abstract

Mathematical decision making theory has been successfully applied to the neuroscience of sensation, behavior, and cognition, for more than fifty years. Classical models rely on the assumption that the environment doesn't change during the period of observation. This assumption has been relaxed in more recent studies of adaptive decision making. We develop new ideal observer – Bayes-optimal – models for this latter setting; and more specifically for the case in which temporal integration of noisy evidence improves choice accuracy. The generative model of the stimulus is a Hidden Markov Model that the ideal observer must filter, and more generally learn. In a first part, we derive and study models tailored to pulsatile evidence with Poisson-distributed timing. We characterize the model parameters that determine choice accuracy, and compare the ideal observer to a finely tuned linear-leak model. We show that the linear model is both more sensitive to parameter perturbation and easier to fit to choice data. In a second part, we derive Bayes-optimal models that learn the change rates of their environment. We do so in several configurations: in discrete time, in continuous time, when more than one change rate must be learned, and for both pulsatile and continuously arriving, drift-diffusion type evidence. We find that such learning models may outperform wrongly tuned known-hazard rate models, but are hard to implement computationally. We conclude that the mathematical study of optimal decision making is crucial for at least three reasons. First, it helps develop an intuition about the various computations required to perform a task. Second, Bayes-optimal models allow benchmarking accuracy and other dependent variables from experiments. Finally, from them, approximate schemes may be built, hopefully taking us one step closer to understanding the human brain.

Contents

1 Introduction **1**

2 Decision making models when change rates are known **11**

2.1 Discrete time decision making models 11

2.1.1 Filtering an N-state Hidden Markov Model 11

2.1.2 Optimal decision making in 2-state environments 14

2.2 Continuous time decision making models 18

2.2.1 Constant 2-state environment 18

2.2.2 Sampled-time approximation 21

2.2.3 Changing 2-state environment with asymmetric rates 22

2.2.4 Changing 2-state environment with symmetric rates 23

2.2.5 Ideal observer models with pulsatile evidence 24

2.3 Summary 28

3 Decision making with pulsatile evidence **30**

3.1 SNR in the Poisson clicks task 30

3.2 Time rescaling in the dynamic clicks task 31

3.3 Effective parameters in the dynamic clicks task 33

3.4 Best linear model and sensitivity analysis 33

3.5 Model fitting . . . 37
3.5.1 The models . . . 38
3.5.2 Bayesian parameter estimation . . . 39
3.5.3 Error in parameter space . . . 41
3.5.4 Deterministic model fits . . . 42
3.5.5 Stochastic model fits . . . 44
3.5.6 Results . . . 48
3.6 Summary . . . 49

4 Learning the change rates in discrete time 50
4.1 Learning a single hazard rate . . . 51
4.1.1 In a 2-state environment . . . 51
4.1.2 In an N-state environment . . . 62
4.2 Learning all transition rates . . . 64
4.3 Summary . . . 72

5 Learning the hazard rate in continuous time 73
5.1 Joint inference on state and hazard rate . . . 73
5.2 Joint inference on state and change-point count . . . 74
5.3 Learning in the dynamic clicks task . . . 79
5.3.1 No-click time window (Case 1) . . . 80
5.3.2 One-click time window (Case 2) . . . 82
5.3.3 Continuum limit . . . 84
5.4 Summary . . . 86

6 Conclusions 87

A Numerical methods for free response protocol 89

B Misattribution noise for pulsatile evidence **90**

Bibliography **92**

List of Figures

1.1 **Illustration of an ambiguous stimulus.** Is this the picture of a sunset or a sunrise? 2

1.2 **Stimulus from the triangles task (image adapted from [23]).** On each trial, a red star appears at a random location on the screen. The location is sampled from one of two 2-dimensional Gaussian densities with similar covariance and means centered at the triangles' locations. The color scale indicates to the subject the height of each density (green = high density; blue = low density). The task consists in choosing the distribution (i.e., the triangle) that generated the star. 7

1.3 **Schematic of the dynamic clicks task.** On each trial, a rat is simultaneously presented with two distinct streams of clicks (one per ear). Each stream has inhomogeneous Poisson statistics with two possible arrival rates, $\lambda_{\text{low}}, \lambda_{\text{high}}$. The transitions between arrival rates are synchronized across streams and are, themselves, governed by an independent Poisson process with rate h, the so-called *hazard rate*. In the figure, the color blue (resp. green) codes for the state $S = 1$ (resp. $S = 0$), which corresponds to a high rate on the right ear and low rate on the left ear (resp. high rate on the left ear and low rate on the right ear). The task is to turn towards the side with higher rate at the end of the trial [34]. 8

2.1 **Single trajectory of the log likelihood ratio y_t, during an artificial trial of the dynamic clicks task.** The decision variable (blue trace) evolves according to equations (2.24). Discontinuous jumps occur at click times in the direction corresponding to the side of the click, and a nonlinear decay to 0 characterizes the periods in-between clicks. Poisson clicks for the right and left streams (black ticks), and environmental state (background color), are depicted above the graph. 28

3.1 **Two effective parameters determine accuracy. A:** Choice accuracy of the ideal observer model from equation (2.24) as a function of rescaled interrogation time T^τ, for two distinct values of $\mathcal{S}^\tau$ (see table 3.1). Two distinct pairs of click rates were used in simulations ($\lambda^\tau_{\text{low}} = 30$ and 60 Hz and $\lambda^\tau_{\text{high}}$ picked to match the appropriate $\mathcal{S}^\tau$ level), yielding the overlaying solid and dashed lines. Keeping T^τ and $\mathcal{S}^\tau$ fixed yields a match in performance for both pairs of click rates. As time evolves during the trial, performance saturates to a maximal value less than 1. **B:** Numerically computed maximal accuracy (color level) of the ideal observer model is constant along the level curves (black lines) of $\mathcal{S}^\tau$ (seen as a function of $(\lambda^\tau_{\text{low}}, \lambda^\tau_{\text{high}})$). We only display the first octant of the $\lambda^\tau_{\text{high}}\lambda^\tau_{\text{low}}$-plane because of the requirement, $0 < \lambda^\tau_{\text{low}} < \lambda^\tau_{\text{high}}$. 34

3.2 **Accuracy of tuned linear model matches the one from the ideal observer. A:** Strong agreement in performance between the nonlinear and linear model (with best γ^*) for a wide range of $\mathcal{S}^\tau$ values. **B:** Best linear discounting rate γ^* for equation (3.3) as a function of $\mathcal{S}^\tau$ (defined in table 3.1). 35

3.3 **Model's sensitivity to discounting parameter perturbation. A:** Performance profile around the optimal discounting parameter for the nonlinear (solid) and linear (dashed) models, presented at several $\mathcal{S}^\tau$-levels (see annotations on curves, or color coding). The linear model is systematically more sensitive to parameter perturbation than its nonlinear counterpart. The relative perturbation c is defined in equations (3.4). **B:** Curvature (absolute value of the second derivative) of the curves from panel A, evaluated at their peak, as a function of $\mathcal{S}^\tau$. The curvature, and hence the sensitivity, of the linear model clearly separates from the curvature of the nonlinear model, for moderate to high $\mathcal{S}^\tau$-values. 36

3.4 **Stabilization of the likelihood estimate yielded by our Monte-Carlo sampling algorithm, as the number of particles increases.** The estimated likelihood (blue curves) is given by the fraction of particles that lands on the correct side of the decision boundary ($y \equiv 0$), in evidence space. The algorithm is applied to the stochastic nonlinear model (3.7), for two distinct trials (columns), two noise values (rows), and two hazard rates (sub-divisions of rows). We used independent particles for distinct points on the x-axis. The red vertical lines indicate the number of particles that we settled on for the other simulations (figures 3.5 and 3.6). 46

3.5 **Validation of our Monte-Carlo sampling algorithm.** For two random trials (columns), two noise values (rows), and two discounting parameter values (sub-divisions of rows), we superimpose the theoretical Gaussian density from section 3.5.5 on the histogram of landing heights of 800 particles. Each particle samples the multiplicative noise applied to clicks height independently, and evolves according to equation (3.6). We observe agreement between the theoretical densities and the histograms. Note: we used our sampling algorithm on the linear model, here, only to be able to compare its end-results to available theoretical densities. 47

3.6 **Relative error of fitting procedures as a function of block size.** Each column (color-coded) corresponds to a specific pair m_f-m_d, where L and NL stand for "linear" and "nonlinear" respectively. Each row (coded by symbol shape) refers to the version of the models used to, both, produce the data and perform the fit. For instance, the panel containing the blue triangles means that we fitted the stochastic linear model to choice data synthetically produced by the stochastic nonlinear model. Error is decreasing in every panel. It decreases faster for deterministic models than for stochastic models, and also faster for the fits of the linear model than for the fits of the nonlinear model. . 48

4.1 **Online inference of the change rate in a dynamic environment (image adapted from [38]). A:** The environment S alternates between states 0 and 1 with transition probabilities η^{01}, η^{10}. We analyze the symmetric case ($h := \eta^{10} = \eta^{01}$) in section 4.1.1 and the asymmetric case $\left(\eta^{01} \neq \eta^{10}\right)$ in section 4.2. The state of the environment determines the sampling distribution of the observations, $f_i(o_n) \coloneqq \Pr\left(o_n \mid S = i\right)$, which we represent as a Gaussian density. **B:** A sample path of the environment (color bar) together with the first ten values of the actual change-point count (a_n) and non-change-point count (b_n). **C:** Evolution of the conditional probabilities, $\Pr\left(h \mid a_n\right)$ (given by Beta distributions), corresponding to the change-point count from panel B, until $n = 100$. The dashed red line indicates the value of h in the simulation. The densities are scaled so that each equals 1 at the mode. 54

4.2 **Graph underlying the transport of joint posterior probability mass across time (image adapted from [38]).** The joint posterior over state and change-point count, $\Pr_n(i, a) \coloneqq \Pr(S_n = i, a_n = a \mid \mathcal{O}_n)$, is propagated along a directed graph according to equation (4.19). A colored node in the graph represents a particular point (S_n, a_n) from the support of the joint posterior. The reader should imagine some probability mass associated to each node in a column, during a specific time of the on-line inference process. At the next time step, the mass is transferred across the edges with the appropriate annotated weights (see equations (4.17) and (4.18) for the definition of these weights). For clarity, only paths corresponding to the initial condition $(S_1, a_1) = (1, 0)$ are shown. 56

4.3 **Learning the hazard rate (image adapted from [38]). A:** Evolution of the posterior over a_n (gray scale). The posterior mean (red) converges to the expected number of change-points $h(n-1)$ (dashed line). **B:** Evolution of the posterior over the change rate h (gray scale). The posterior mean (red) converges to the true value (dashed line) and the variance diminishes with the number of observations. 60

4.4 **Performance of the inference algorithm (image adapted from [38]). A:** Accuracy under the interrogation paradigm measured as the percentage of correct responses at the interrogation time. Here and in the next panel $h = 0.05$, and SNR= 1. The black curve represents the performance of an ideal observer who infers the change rate from measurements. The green curves represent the performance of observers that assume a fixed change rate (0.3, 0.15, 0.05, 0.03 from darker to lighter, see equation (2.9)). The solid green line corresponds to an observer who assumes the true rate, dashed lines to erroneous rates. **B:** The green curve represents the performance at interrogation time $n = 300$ of an observer that assumes a fixed change rate (x-axis). The red star marks the maximum of this curve, corresponding to the true change rate $h = 0.05$. The horizontal black lines represent the accuracy at times $n \in \{40, 100, 200, 300\}$ (from bottom to top) of the observer that learns the change rate. **C:** The accuracy as a function of the average threshold hitting time in the free response protocol. Here $h = 0.1$, and SNR=0.75. See appendix A for details on numerical simulations. See also figure 3 in [49]. 61

4.5 **Evidence accumulation and change rates inference in a 2-state asymmetric system (image adapted from [38]). A:** Sample path (color bar, top) of the environment between time steps 70 and 80 (same simulation as in panels C-E) with corresponding observations (blue dots), and log posterior odds ratio (black step function). Here and in panels C-E, $\left(\eta^{01}, \eta^{10}\right) = (0.2, 0.1)$, SNR= 1.4. **B:** The number of allowable change-point count matrices as a function of observation number, n, for $M = 2$ (blue circles), and $M = 3$ (blue triangles). **C-E:** Color plots (gray scale) of the joint density, $\Pr\left(\eta^{01}, \eta^{10} \mid \mathcal{O}_n\right)$, with mean value (red star) approaching the true transition rates (green circle), as number of observations increases. 71

List of Tables

1.1 **The four combinations of answered class ($S_{\mathbf{ans}}$) and true class (S) in a 2AFC task.** The state of the environment S may take on one of two values, 0 or 1, and the observer must identify this state based on noisy evidence. The four combinations may be classified into correct and incorrect trials. 4

3.1 **Parameters from the ideal observer model for the dynamic clicks task, expressed in two different time units.** The time variable t may represent seconds, while the time variable τ is such that one unit of τ equals $1/h$ seconds. Note that SNR and κ are not affected by this transformation. 32

5.1 **Definition of the constants in front of the delta-jumps in equation** (5.34)**.** 85

Chapter 1

Introduction

All living beings must cope with changing environments to survive. From the regulatory mechanisms within a single cell to the stock brokers on Wall street who try to maximize their return, Life is all about coping with uncertainty and changes in the environment. A modern paradigm to think about animal[1] behavior is that of *decision making*. Every time a biological system selects an action among a set of alternatives, we say that it makes a decision. Thus, the type of decision that we study here is more general than the layman's connotation for the term. A neuron *decides* to fire or not, given its synaptic inputs [17]; a bee colony *decides* to move its hive to one of N potential locations, based on the information collected by scout bees [45]; and a teenager *decides* to apply to specific colleges at the end of Senior year, based on a complex social and psychological context [9]. In the present work, we are interested in uncovering the mechanisms and biological principles that allow living organisms to make decisions in changing environments.

Decision making theory[2] has met such scientific success in the past century that it today pervades most branches of biology [5], finance [14] and engineering [28]. A great example of its power is the invention, by Alan Turing and colleagues, of a decision making algorithm that broke the code of the German Enigma machine during

[1] In this work, the word "animal" also refers to humans.

[2] Also simply called decision theory.

Figure 1.1: **Illustration of an ambiguous stimulus.** Is this the picture of a sunset or a sunrise?

World War II [24]. So, what exactly is decision making theory[3]? **It is a mathematically principled way of using all available information to determine the odds of different outcomes, and then using these odds to determine the best choice.** Closely related to game theory, it postulates and leverages a statistical structure of the world in which each possible action is associated with some probability of reward and cost. Importantly, the decision maker – interchangeably called the *agent* or the *observer* for the remainder of this text – makes at least one observation of the state of the world before deciding on an action. The observations are *noisy* insofar as they only provide ambiguous information about the state of the world. It is the pair ("state of the world", "action taken") which ultimately determines the payoff. For example, I could ask you to decide whether the picture from figure 1.1 represents that of a sunset or a sunrise. I could further decide to reward/penalize you[4] in the following way:

[3]Some researchers might encompass a wider field than the one we are describing here with this locution, reserving the qualifier "normative decision making theory" for our work [39]. We totally agree with these further distinctions. Our present purpose is *not* to define the boundaries of the field, but rather to introduce our topic in a simple way.

[4]To play this mental game, let's assume that you want to play and trust me with the truth.

- You pay me \$10 if you miss a sunrise.
- You pay me \$1 if you miss a sunset.
- You earn \$5 if you correctly guess a sunrise.
- You earn \$5 if you correctly guess a sunset.

Here, the state of the world is the truth, and the observation consists of you looking at the picture. The decision is your answer. Clearly, with no prior information about the picture, and unless you know of a way of distinguishing between pictures of sunrise and sunset, you should answer "sunrise", irrespective of what the truth is, since you lose more money by missing a sunrise. But what if you knew where the picture was taken? If, somehow, you found some clue that it was taken in France, you could become much more confident that the coast is facing west, and therefore that it is a sunset picture! Decision making theory formalizes this procedure of balancing out the beliefs, the new information and the reward structure, in some "best possible way".

Two set-ups must be distinguished: In one, the *environment* – i.e., the state of the world – is *static* during the observation period (as in our picture example); in the other, it is *dynamic*. Signal Detection Theory (SDT) [26, 19] lays the mathematical grounds for making optimal decisions in the former setting. We will present its main results before introducing our work, which focuses on the latter setting.

In SDT, a presented stimulus x must be categorized as pertaining to one of two classes of stimuli: "noise alone" (n) or "noise with added signal" (sn). Mathematically, these classes are represented by two conditional probability distributions for x: $\Pr\Big(x \mid \mathrm{n}\Big)$ and $\Pr\Big(x \mid \mathrm{sn}\Big)$. On a given trial, x is sampled from one of these distributions and presented to a *detector* [19] whose task is to answer "Yes" if x was generated from $\Pr\Big(x \mid \mathrm{sn}\Big)$ and "No" otherwise. This task becomes nontrivial when the support of the two distributions overlap since any value of x lying in the intersection of the supports may be generated by either class with nonzero probability. We want to stress from the outset that this mathematical setting need not be restricted to the SDT nomenclature. From a more general perspective, the stimulus is an observation o, which is sampled from one of two distributions, $\Pr\Big(o \mid S = 0\Big)$ and $\Pr\Big(o \mid S = 1\Big)$. We call the distribution that generates the sample – or, equivalently, the random variable S – the *state of the environment*. This decision making paradigm is referred to as a 2-Alternative Forced Choice (2AFC) task.

Since the observer's answer S_{ans} depends on the state and the stochastic observation, it is itself a random variable. Table 1.1 illustrates the four possible combina-

	$S = 0$	$S = 1$
$S_{\text{ans}} = 0$	correct	incorrect
$S_{\text{ans}} = 1$	incorrect	correct

Table 1.1: **The four combinations of answered class (S_{ans}) and true class (S) in a 2AFC task.** The state of the environment S may take on one of two values, 0 or 1, and the observer must identify this state based on noisy evidence. The four combinations may be classified into correct and incorrect trials.

tions of (S_{ans}, S) on any given trial. Two of these combinations correspond to correct trials, and the remaining two to incorrect trials. Typically, a reward structure assigns a payoff value $V(S_{\text{ans}}, S)$ to each combination, which defines a *value function*. Equivalently, many authors change the sign of the value function to define a *loss function* [39]. From there, three decision goals are commonly studied [19]:

1. Maximize the expected value (equivalently minimize the expected loss) across trials.

2. Maximize the probability of being correct on any given trial.

3. Maximize the probability of correctly classifying one class for a given tolerance level on the probability of wrongly classifying the other class[5]. For instance, maximize $\Pr\left(S_{\text{ans}} = 1 \mid S = 1\right)$ under the constraint that $\Pr\left(S_{\text{ans}} = 1 \mid S = 0\right) < \alpha$, for some tolerance level $\alpha > 0$.

Once a goal has been set, an *optimal decision making strategy*, also called an *optimal policy*, is a mapping $x \mapsto S_{\text{ans}}$ that meets the goal. It has been known for a long time that the goals presented above overlap under certain simplification assumptions [19]. For example, as long as

$$V(1,1) - V(0,1) = V(0,0) - V(1,0) \tag{1.1}$$

the optimal strategies for goals 1 and 2 are the same. Also, there always is a tolerance level α in goal 3 that renders the corresponding optimal strategy identical to that from goal 2.

[5]This is the *Neyman-Pearson* criterion of optimality [31].

In the present work, we almost exclusively study decision making models for 2AFC tasks in which equation (1.1) holds. Furthermore, we always assume goal number 2 above. Given these assumptions, it is known (see equation (1.19) in [19] for instance) that the optimal strategy is to base one's decision on the comparison of the so-called *likelihood ratio*,

$$\frac{\Pr\Big(o \mid S=1\Big)}{\Pr\Big(o \mid S=0\Big)}$$

to the *decision threshold*,

$$\frac{\pi\left(S=0\right)}{\pi\left(S=1\right)}$$

where $\pi(S)$ represents the *prior odds* of each environmental state. If the likelihood ratio is greater (resp. smaller) than the threshold, the observer should choose state 1 (resp. 0). In case of equality, a uniform sample from $\{0,1\}$ should be selected. This decision rule is a direct application of Bayes' rule in order to decide which of the two alternatives $\{0,1\}$ has highest *posterior probability*:

$$\frac{\Pr\Big(o \mid S=1\Big)}{\Pr\Big(o \mid S=0\Big)} \gtrless \frac{\pi\left(S=0\right)}{\pi\left(S=1\right)} \Leftrightarrow 1 \lesseqgtr \frac{\Pr\Big(o \mid S=1\Big)\,\pi\left(S=1\right)}{\Pr\Big(o \mid S=0\Big)\,\pi\left(S=0\right)} = \frac{\Pr\Big(S=1 \mid o\Big)}{\Pr\Big(S=0 \mid o\Big)}$$

As such, this decision strategy is sometimes called *Bayes-optimal*. For the remainder of this work, we call *ideal observer* any decision making model implementing this strategy.

In the SDT framework, nothing prevents the observation from being a vector of n independent[6] and identically distributed (i.i.d) random variables, $o = (o_1, \ldots, o_n)$. In this case, their joint likelihood factors into a product of n individual likelihoods, and this is exactly the setting of the Sequential Probability Ratio Test (SPRT) that we derive in chapter 2. Thus, evidence from multiple noisy samples can be combined in order to improve the probability of a correct decision. In a physical system, these observations can be spread across space or time. When spread across time, the corresponding theory of evidence integration is called *sequential analysis* [50, 24]. As independent samples are gathered sequentially in time, the observer iteratively combines the accumulated information into a time-dependent log posterior odds ratio,

[6] When conditioned on the environment.

which is nothing else than the log likelihood ratio scaled by the log prior odds ratio:

$$\begin{aligned} y_n &:= \log \frac{\Pr\Big(S=1 \;\Big|\; o\Big)}{\Pr\Big(S=0 \;\Big|\; o\Big)} \\ &= \log \frac{\Pr\Big(o \;\Big|\; S=1\Big)}{\Pr\Big(o \;\Big|\; S=0\Big)} + \log \frac{\pi(S=1)}{\pi(S=0)} \\ &= \log \frac{\Pr\Big(o_n \;\Big|\; S=1\Big)}{\Pr\Big(o_n \;\Big|\; S=0\Big)} + y_{n-1} + \log \frac{\pi(S=1)}{\pi(S=0)} \end{aligned}$$

Two experimental paradigms are commonly used in neuroscience. In the *interrogation protocol*, the length of the observation time window is preset by the experimenter. In the *free response protocol*, the subject is free to respond at any time they want. The Bayes-optimal strategy for performing the task, in both protocols, relies on comparing the log posterior odds ratio to one or two decision thresholds. In the interrogation protocol, the single threshold is $y \equiv 0$. Denoting interrogation time by N, the ideal observer answers "0" if $y_N < 0$, "1" if $y_N > 0$, and selects either option with equal probability if $y_N = 0$. In the free response protocol, we assume prior confidence thresholds, $y \equiv \pm y_{\text{threshold}}$, such that a decision is made as soon as the accumulated evidence reaches them. If $y_{\text{threshold}}$ is reached first, the ideal observer selects choice "1", and if $-y_{\text{threshold}}$ is reached first, it selects choice "0".

In the natural world, the environment is seldom constant during the observation period that leads to a decision. We study optimal evidence integration in dynamic environments. Intuitively, one may reason that the older the gathered evidence gets, the less relevant it becomes for identifying the present state of the environment. This is indeed what normative theory prescribes: The differential equations that govern the evolution of the log posterior odds ratio contain a leak term [17, 23, 49].

In behavioral and cognitive neuroscience, recent experiments have been designed to study adaptive perceptual[7] decision making. Our theoretical work is specifically tailored to understand three of such experiments. In the last decade, the lab of Josh Gold [2] at the University of Pennsylvania has been implementing several variants of the *triangles task* and the *dots reversal task* [23, 22]. More recently, the lab of Carlos Brody [1] at Princeton University has been studying the *dynamic clicks task* [34]. We briefly present these tasks in their standard form below.

[7]The adjective "perceptual" refers to the fact that the animal must base its decision on a perceived sensory stimulus.

In the triangles task [23], on each trial, a red star appears at a random location

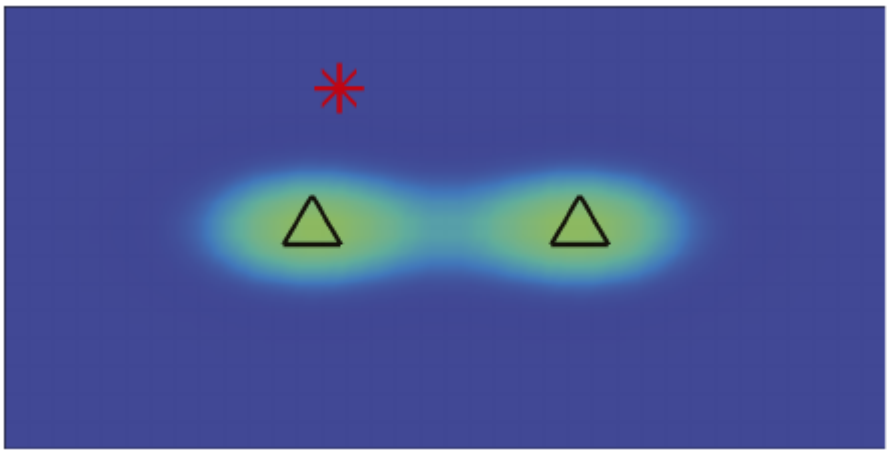

Figure 1.2: **Stimulus from the triangles task (image adapted from [23]).** On each trial, a red star appears at a random location on the screen. The location is sampled from one of two 2-dimensional Gaussian densities with similar covariance and means centered at the triangles' locations. The color scale indicates to the subject the height of each density (green = high density; blue = low density). The task consists in choosing the distribution (i.e., the triangle) that generated the star.

on the screen (see figure 1.2). The location is sampled from one of two 2-dimensional Gaussian densities with similar covariance and means centered at the triangles' locations. The color scale indicates to the subject the height – and therefore the spread – of each density (green = high density; blue = low density). The task consists in choosing the distribution (i.e., the triangle) that generated the star. Trials are presented in blocks of constant task parameters. Across any two consecutive trials within a block, there is some probability h that the correct choice switches side. The h-variable is termed the *hazard rate*. It is set by the experimenter to be constant within blocks, and to change across blocks, spanning the range $0.05 - 0.95$. Since a single star and choice occur on each trial, this task is better modeled in discrete time.

The dots reversal task [23] is both designed for humans and monkeys. It is an extension of the traditional *random moving dots task*, the latter of which has lead to a vast amount of research [44, 11, 42, 25]. In the dots reversal task, the subject is presented with a cloud of randomly moving dots [3] during a period of time that lasts between 5 and 10 seconds. A fraction of the dots moves coherently in one of two directions (left or right), while the remaining dots move randomly and independently

of each other. Importantly, a *hazard rate* h, expressed in Hz, controls the frequency at which the coherent direction of motion switches within a trial. These switches occur at random times according to a Poisson process with rate h. The task consists in making an eye saccade[8], at the end of the trial, in the direction in which the coherent dots were last perceived to move. Each subject undergoes blocks of trials with constant hazard rate. The two values of h tested in [23] were 0.1Hz and 2Hz.

The *dynamic clicks task* [34] is an extension of the *Poisson clicks task* [12], which has been performed on humans and rats. In the auditory version of the dynamic clicks task (see figure 1.3), two streams of auditory clicks are simulatneously delivered to a subject's ears during each trial (one stream per ear). Each click train, lasting

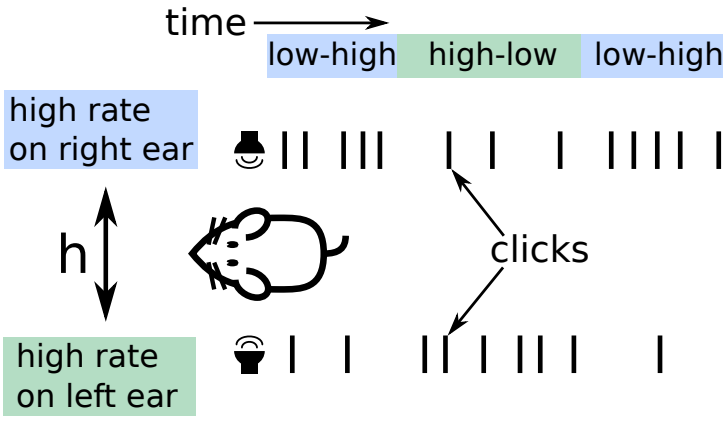

Figure 1.3: **Schematic of the dynamic clicks task.** On each trial, a rat is simultaneously presented with two distinct streams of clicks (one per ear). Each stream has inhomogeneous Poisson statistics with two possible arrival rates, $\lambda_{\text{low}}, \lambda_{\text{high}}$. The transitions between arrival rates are synchronized across streams and are, themselves, governed by an independent Poisson process with rate h, the so-called *hazard rate*. In the figure, the color blue (resp. green) codes for the state $S = 1$ (resp. $S = 0$), which corresponds to a high rate on the right ear and low rate on the left ear (resp. high rate on the left ear and low rate on the right ear). The task is to turn towards the side with higher rate at the end of the trial [34].

between 0.5 and 2 seconds, is generated by an inhomogeneous Poisson process with stepwise constant arrival rate that is restricted to two values (λ_{low} vs. λ_{high}). The two streams are coupled in the sense that they have same transition times. That is, whenever the arrival rate of one stream undergoes a sudden change $\lambda_{\text{low}} \to \lambda_{\text{high}}$, the other stream undergoes the opposite change $\lambda_{\text{high}} \to \lambda_{\text{low}}$. Both ears always receive

[8]Eye-tracking devices allow the experimenter to record the subject's choice.

clicks at distinct arrival rates. The task consists in choosing the stream with highest arrival rate at the end of the trial.

In all three experiments presented above, the true state of the environment (the one that is rewarded) is in one of two states and alternates during the trial in a history-independent way. The subject receives sequential noisy samples from the hidden state until interrogation time, at which point they are required to identify the state in which the trial ended.

In neuroscience, initial models of optimal decision making in changing environments assumed that the statistics of the changes were known to the observer. This assumption was a reasonable first step, given that these statistics are key to maximizing accuracy: Knowing the average number of changes that will occur in a trial allows, in a Markovian setting, to integrate evidence with the appropriate timescale. In this work, we will both analyze existing models and derive new ones in an attempt to relax this assumption.

In chapter 2, we re-derive established ideal observer models from the literature. Their unifying characteristic is that they all assume the change rates of the environment to be known. Their derivation is important for the present work, because it involves key concepts and techniques that we use throughout the dissertation. In particular, we introduce there the Sequential Probability Ratio Test (SPRT), from which we derive the Drift-Diffusion Model (DDM) as its *continuum limit*.

In chapter 3, we study in detail the decision model from the previous chapter that concerns pulsatile evidence – this is the model for the dynamic clicks task. We show that, via an appropriate time rescaling, only two effective task parameters govern the accuracy of the ideal observer on the dynamic clicks task. We then explore how the best *linear* discounting model of evidence accumulation compares to the ideal observer model. Our comparison operates in terms of sensitivity to parameter perturbation, and of ease of model fitting.

Chapters 4 and 5 develop ideal observer models that *learn* the change rates of their environment. Chapter 4 studies discrete time environments, and chapter 5, continuous time ones. By learning models, we mean decision making models that compute and iteratively update a posterior distribution over the change rates, as observations are gathered. Because we assume that the goal of our model is to maximize choice accuracy, computing a posterior over change rates alone is not sufficient. We therefore leverage the notion of *joint posterior* in two separate contexts. On the one hand, we derive models that update a joint posterior over environmental state and hazard rate. On the other hand, we focus on a joint posterior over environmental

state and change-point count. All such models, and especially the ones in continuous time, are hard to analyze. We hope that, in the future, approximate schemes can be developed on their basis.

Chapter 2

Decision making models when change rates are known

This chapter lays down the theoretical foundations for modeling sequential decision making. Although most of its content was already established prior to our work, we present it here in a unified manner, as an introduction to the concepts and techniques that will be used later. Throughout the chapter, the mathematical problem of optimally deciding which of M decisions is the correct one, is framed as a filtering problem on a Hidden Markov Model (HMM). The change rates of the hidden chain are assumed known for now. The question of learning them will be addressed in chapters 4 and 5.

A word on notation: In this work, we do not include 0 in the set of natural integers $\mathbb{N}$.

2.1 Discrete time decision making models

2.1.1 Filtering an N-state Hidden Markov Model

The experimental set-up in which decision making is studied lends itself particularly well to the theoretical framework of HMMs. Two key concepts are at play: the fact that a *hidden* signal must be used by the observer in order to solve the task, and the fact that this signal is *Markovian*. Indeed, experimental tasks of perceptual decision

making can be framed as a signal decoding problem. The variable that is relevant to the task[1] (the signal) is not directly observed. Instead, the observer must rely on the conditional probabilities that relate the observed stimulus (the corrupted signal) to the hidden variable.

Let $\{S_n\}_{n\in\mathbb{N}}$ be a Markov chain with state space $\Lambda \coloneqq \{1,\dots,M\}$ and transition matrix $\left(\eta^{ij}\right)_{(i,j)\in\Lambda^2}$. Each transition rate η^{ij} represents the probability of the chain transitioning from state j to i in a single time step (here, time is discrete and starts at $n=1$). Let $\{o_n\}_{n\in\mathbb{N}}$ be another discrete time process with state space $\Xi \subseteq \mathbb{R}$, representing partial observations of the state chain $\{S_n\}_{n\in\mathbb{N}}$. More precisely, the following two conditions hold.

1. For each state $i \in \Lambda$, there is a probability density function[2] $f_i : \Xi \to \mathbb{R}$, such that for all $n \in \mathbb{N}, o \in \Xi$, $f_i(o) \coloneqq \Pr\left(o_n = o \;\middle|\; S_n = i\right)$.

2. The random variables $\{o_n\}_{n\in\mathbb{N}}$ are conditionally independent on the state. That is, for any finite subset $\{n_1,\dots,n_N\} \subset \mathbb{N}$,

$$\Pr\left(o_{n_1},\dots,o_{n_N} \;\middle|\; S_{n_1},\dots,S_{n_N}\right) = \prod_{k=1}^{N} f_{S_{n_k}}(o_{n_k}) \tag{2.1}$$

With $\pi : \Lambda \to [0,1]$, a prior distribution over the states, the tuple,

$$\left(\{S_n\}_{n\in\mathbb{N}},\ \pi,\ \{o_n\}_{n\in\mathbb{N}},\ \{f_i\}_{i\in\Lambda}\right)$$

constitutes a Hidden Markov Model[3]. Calling $\mathcal{O}_N \coloneqq \{o_1,\dots,o_N\}$ a particular realization of the first N steps of the observation process, *filtering* the HMM at time N means computing the conditional probabilities $\Pr\left(S_N = i \;\middle|\; \mathcal{O}_N\right)$, for all $i \in \Lambda$. In this section, we present the solution to this problem when the transition rates $\left(\eta^{ij}\right)_{(i,j)\in\Lambda^2}$ are known.

Disclaimer: Here, and throughout the dissertation, we make the assumption that the conditional probabilities and density functions mentioned always exist. This necessarily happens if we assume that the state chain $\{S_n\}_{n\in\mathbb{N}}$ is *ergodic* – every state

[1] The one that determines the reward.

[2] If Ξ is discrete, then the f_i's are probability mass functions.

[3] See [37] for a tutorial on HMMs and [13] for an in-depth exposition.

in Λ is positive recurrent and aperiodic – and that the observation noise densities $\{f_i\}_{i\in\Lambda}$ *vanish nowhere*.

As we will soon demonstrate, the posterior probability of state i at time step $n \in \mathbb{N}$ obeys the recursive equation,

$$\Pr\Big(S_n = i \;\Big|\; \mathcal{O}_n\Big) \propto f_i(o_n) \sum_{j=1}^{M} \eta^{ij} \cdot \Pr\Big(S_{n-1} = j \;\Big|\; \mathcal{O}_{n-1}\Big) \qquad \forall n > 1 \tag{2.2a}$$

$$\Pr\Big(S_1 = i \;\Big|\; \mathcal{O}_1\Big) = \frac{1}{\Pr\left(\mathcal{O}_1\right)} f_i(o_1)\pi(S = i) \tag{2.2b}$$

where the proportionality constant in (2.2a) only depends on the observation stream $\mathcal{O}_n$. With N representing the end of observation time in the interrogation protocol, the ideal observer chooses the state $\hat{\imath}$ which displays highest posterior probability:

$$\hat{\imath} \coloneqq \mathrm{argmax}_{i\in\Lambda} \left\{ \Pr\Big(S_N = i \;\Big|\; \mathcal{O}_n\Big) \right\} \tag{2.3}$$

Because equation (2.2a) is recursive[4], it is termed an *on-line inference algorithm*. Alternative algorithms[5] using parts or all of the observation history to compute the posterior over the state variable are termed *off-line* algorithms. This latter type of algorithm is useful to infer a state value that occurred before interrogation time[6]. That is, an off-line algorithm is required to compute $\Pr\Big(S_k = i \;\Big|\; \mathcal{O}_N\Big)$ for $k < N$. Filtering and smoothing HMMs represents a whole field of study of its own, with numerous applications in engineering, finance and artificial intelligence [13].

However, as equation (2.2a) demonstrates, on-line inference is possible when the goal is to infer the present state of the environment.

We now set out to derive equation (2.2). The initial conditions (2.2b) are a literal application of Bayes' rule. As for (2.2a), we start by using the law of total probability as follows. For any $n \in \mathbb{N}_{\geq 2}$ and $i \in \Lambda$:

$$\Pr\Big(S_n = i \;\Big|\; \mathcal{O}_n\Big) = \sum_{j\in\Lambda} \Pr\Big(S_n = i, S_{n-1} = j \;\Big|\; \mathcal{O}_n\Big) \tag{2.4}$$

[4]By this we mean that the posterior at time step n only depends on the current observation and on the posterior at time step $n-1$.

[5]Notable ones are the Forward-Backward algorithm [6] and the Expectation-Maximization algorithm [16].

[6]This is referred to as *smoothing* in the signal processing literature.

Next, Bayes' rule applied to the summand of the right-hand side above yields,

$$\Pr\left(S_n = i \mid \mathcal{O}_n\right) \propto \sum_{j \in \Lambda} \Pr\left(\mathcal{O}_n \mid S_n = i, S_{n-1} = j\right) \Pr\left(S_n = i, S_{n-1} = j\right) \tag{2.5}$$

with proportionality constant $1/\Pr\left(\mathcal{O}_n\right)$. From here, we note two things. First, since the observations are conditionally independent on the states[7], the following holds:

$$\Pr\left(\mathcal{O}_n \mid S_n = i, S_{n-1} = j\right) = \Pr\left(o_n \mid S_n = i\right) \Pr\left(\mathcal{O}_{n-1} \mid S_{n-1} = j\right)$$

Second, from basic probability theory, we have:

$$\Pr\left(S_n = i, S_{n-1} = j\right) = \Pr\left(S_n = i \mid S_{n-1} = j\right) \Pr\left(S_{n-1} = j\right)$$

Combining these two manipulations on the summand of equation (2.4) gives the new summand:

$$\Pr\left(o_n \mid S_n = i\right) \Pr\left(\mathcal{O}_{n-1} \mid S_{n-1} = j\right) \Pr\left(S_n = i \mid S_{n-1} = j\right) \Pr\left(S_{n-1} = j\right)$$

Noting that,

$$\Pr\left(\mathcal{O}_{n-1} \mid S_{n-1} = j\right) \Pr\left(S_{n-1} = j\right) = \Pr\left(S_{n-1} = j \mid \mathcal{O}_{n-1}\right) \Pr\left(\mathcal{O}_{n-1}\right)$$

and remembering our notation,

$$f_i(o_n) \coloneqq \Pr\left(o_n \mid S_n = i\right) \qquad \eta^{ij} \coloneqq \Pr\left(S_n = i \mid S_{n-1} = j\right)$$

we are finally able to recover the posterior described by equation (2.2a), with proportionality constant $\Pr\left(\mathcal{O}_{n-1}\right) / \Pr\left(\mathcal{O}_n\right)$.

2.1.2 Optimal decision making in 2-state environments

Most experimental tasks investigating perceptual decision making constrain the environment to two states. Such tasks are termed Two-Alternative Forced Choice (2AFC) tasks in the behavioral science literature. As we will see, the mathematical models for optimal decision making greatly simplify in this setting. The key simplification point lies in the fact that when $\Lambda = \{0, 1\}$, knowing the posterior

[7] Recall equation (2.1).

over one state automatically yields the posterior over the complementary state, as $\Pr\left(S_n = i \mid \mathcal{O}_n\right) + \Pr\left(S_n = 1 - i \mid \mathcal{O}_n\right) = 1$. Furthermore, since an ideal observer only needs to select the state with highest posterior (equation (2.3)), comparing the ratio of posteriors $\Pr\left(S_n = i \mid \mathcal{O}_n\right) / \Pr\left(S_n = 1 - i \mid \mathcal{O}_n\right)$ to one is a sufficient strategy. It is indeed the basis for the algorithms presented below. Since all the models from this section are derived from equation (2.2), we made the choice of presenting them in decreasing order of complexity.

Remark: for notational convenience, when the environment only admits two states, we set $\Lambda \coloneqq \{0, 1\}$ instead of $\Lambda = \{1, 2\}$. As a consequence, the transition rate from state 0 to 1 is written η^{10}, and similarly for the remaining three rates.

Changing environment with asymmetric change rates

As explained above, the strategy of an ideal observer is to compare the ratio of posterior odds to 1 at interrogation time. Without loss of generality[8], we consider the ratio at time $n \in \mathbb{N}$ to be:

$$R_n \coloneqq \frac{\Pr\left(S_n = 1 \mid \mathcal{O}_n\right)}{\Pr\left(S_n = 0 \mid \mathcal{O}_n\right)}$$

If at interrogation time N, $R_N > 1$, then state 1 is more probable than 0, given the observer's prior belief π and the specific history of observations $\mathcal{O}_N$.

Using equation (2.2), and writing $R_0 \coloneqq \pi(S = 1)/\pi(S = 0)$ for the prior odds ratio, we have:

$$R_n = \frac{f_1(o_n)}{f_0(o_n)} \cdot \frac{\eta^{10} \cdot \Pr\left(S_{n-1} = 0 \mid \mathcal{O}_{n-1}\right) + \eta^{11} \cdot \Pr\left(S_{n-1} = 1 \mid \mathcal{O}_{n-1}\right)}{\eta^{01} \cdot \Pr\left(S_{n-1} = 1 \mid \mathcal{O}_{n-1}\right) + \eta^{00} \cdot \Pr\left(S_{n-1} = 0 \mid \mathcal{O}_{n-1}\right)} \qquad \forall n > 1$$

$$R_1 = \frac{f_1(o_n)}{f_0(o_n)} R_0$$

[8] By this, we mean that equivalent decision making models can be based on the inverse of our posterior odds ratio, $1/R_n$.

We may rearrange the right-hand side of the above equation to reveal the term R_{n-1}, and replace η^{ii} by $1 - \eta^{1-i,i}$ for $i \in \Lambda$, to obtain the new equation[9]:

$$R_n = \frac{f_1(o_n)}{f_0(o_n)} \cdot \frac{\eta^{10} + \left(1 - \eta^{01}\right) \cdot R_{n-1}}{\eta^{01} \cdot R_{n-1} + (1 - \eta^{10})} \qquad \forall n > 1 \tag{2.6a}$$

$$R_1 = \frac{f_1(o_n)}{f_0(o_n)} R_0 \tag{2.6b}$$

At this point, it is common practice to study the log posterior odds ratio, which is the natural logarithm of the previous ratio: $y_n \coloneqq \log R_n$. As we will show in later sections, this transformation is useful when taking continuum limits and more stable for numerical simulations. It is also more readily interpretable as neural population activity [24]. The log posterior odds ratio is called the *decision variable* in decision making theory as it is the variable that standard ideal observer models use to make their decision.

Another quantity of importance needs to be explained, before pursuing our calculations. The ratio $f_1(o_n)/f_0(o_n)$ is termed the *likelihood ratio*. It quantifies the relative way in which the current observation o_n favors each environmental state. If an observer with no prior belief about the state of the environment were to use the single observation o_n to decide which of the two states 0 or 1 were more likely, the optimal procedure would be to compare the likelihood ratio to one. A ratio greater than one favors state 1 insofar as it indicates that this state is more likely to give rise to o_n than the opposite state. In a symmetric fashion, a ratio smaller than one favors state 0. A ratio exactly equal to 1 indicates that the current observation is equally likely to have come from either state. After taking the natural logarithm, as we are about to do, the likelihood ratio becomes what is commonly called – and with no surprise! – the *log likelihood ratio*. We denote it by:

$$I_n \coloneqq \log \frac{f_1(o_n)}{f_0(o_n)} \qquad \forall n \in \mathbb{N} \tag{2.7}$$

The letter I is a reference to the word *information* as, informally, I_n represents a piece of information provided by observation o_n about the correct decision.

Returning to our derivation of the ideal observer model, we use the relation

[9] This is equivalent to equation (3.2) in [49].

$R_n = \exp(y_n)$ to rewrite equation (2.6) as:

$$y_n = I_n + \log \frac{\eta^{10} + \left(1 - \eta^{01}\right) \cdot \exp(y_{n-1})}{\eta^{01} \cdot \exp(y_{n-1}) + (1 - \eta^{10})} \qquad \forall n > 1$$
$$y_1 = I_1 + y_0$$

Finally, adding and subtracting $y_{n-1} = \log R_{n-1}$ to the right-hand side gives us the long sought-for recursive equation[10] for the decision variable:

$$y_n = I_n + y_{n-1} + \log \frac{\eta^{10} \cdot \exp(-y_{n-1}) + \left(1 - \eta^{01}\right)}{\eta^{01} \cdot \exp(y_{n-1}) + (1 - \eta^{10})} \qquad \forall n > 1 \tag{2.8a}$$
$$y_1 = I_1 + y_0 \tag{2.8b}$$

As time evolves, the decision variable y_n sequentially accrues pieces of evidence, I_n, about the true state of the environment, while discounting old evidence in a change rates dependent way (third term in the right-hand side of (2.8a)). The log prior odds ratio, $y_0 = \log\big(\pi(S = 1)/\pi(S = 0)\big)$, represents the observer's prior belief about the relative likelihoods of each state. An a priori unbiased belief amounts to $y_0 = 0$. Lastly, instead of comparing R_N to 1, the decision making algorithm now needs to compare y_N to 0; if $y_N > 0$, the ideal observer chooses state 1, and 0 otherwise. In the event that $y_N = 0$, the ideal observer possesses equal evidence in favor of either alternative. A probabilistic decision rule of the form, "*choose 0 with probability 1/2; and 1 otherwise*", is therefore the best strategy to maximize the probability of being correct across independent trials, in this case.

Changing environment with symmetric change rates

When the change rates of the 2-state environment are symmetric, we write $h := \eta^{10} = \eta^{01}$, and as a result, the optimal decision making algorithm from equation (2.8) becomes[11]:

$$y_n = I_n + y_{n-1} + \log \frac{h \cdot \exp(-y_{n-1}) + (1 - h)}{h \cdot \exp(y_{n-1}) + (1 - h)} \qquad \forall n > 1 \tag{2.9a}$$
$$y_1 = I_1 + y_0 \tag{2.9b}$$

As h is increased from 0 to 0.5, more and more evidence is discounted. At $h = 0.5$, $y_n = I_n$. Then, for $h > 0.5$, y_n will tend to switch sign on each observation.

[10] This last equation corresponds to equation (B.1) in [49].
[11] This last equation is equivalent to equation (2) in [23].

Constant environment

The case of a constant environment can be seen as a HMM in which the hazard rate is trivially 0. So, setting $h = 0$ in equation (2.9) yields the following ideal observer model:

$$y_n = I_n + y_{n-1} \qquad \forall n \in \mathbb{N} \tag{2.10}$$

Note how the discounting terms from equations (2.8) and (2.9) have vanished. The reason is that perfect integration of evidence is the optimal decision making strategy when one *knows* that the environment is constant.

In spite of our above presentation of the decision making models in decreasing order of complexity, it must be noted that they were historically discovered and studied in the reverse order. Thus, equation (2.10) was devised by Alan Turing during the second World War to break the code of the German Enigma cryptography machine [24]. Later, this algorithm was termed the *Sequential Probability Ratio Test* by Abraham Wald [50] and spurred an abundance of research in behavioral and cognitive psychology [40, 25].

2.2 Continuous time decision making models

The study of continuous time models of decision making poses the challenge of heavier, more advanced, mathematical machinery than their discrete time counterparts. On the other hand, techniques from the analysis of continuous time stochastic processes are often much easier to apply in order to track the distribution of the decision variable over time. It is ultimately up to the modeler to find their trade-off between these two types of models.

2.2.1 Constant 2-state environment

The continuous time model presented in this section is the *Drift-Diffusion Model* [10, 8]. The decision variable y_t is now a stochastic process that evolves in continuous time ($t \in [0, \infty)$) according to a Stochastic Differential Equation (SDE) of the following form:

$$dy_t = m \cdot dt + D \cdot dW_t \tag{2.11}$$

The initial condition of the above equation is usually deterministic: $y_0 = 0$. The constant terms m and D are called the *drift* and *diffusion* coefficients, respectively.

The general idea behind the DDM is similar as for the SPRT. Sensory evidence is gathered incrementally into the decision variable y_t. The true state of the environment, which is constant within a trial in the present setting, determines the drift rate m. The noisy aspect of the evidence is reflected in the diffusion term.

Although the DDM can be used without any reference to the SPRT, drawing a formal correspondence between the two enriches one's understanding of both models. As was shown in section 2.1.2, the SPRT is Bayes-optimal. Thus, deriving the DDM as the continuum limit of the SPRT makes the DDM inherit this optimality property.

DDM as continuum limit of SPRT

The continuum limit makes use of Donsker's invariance principle, so-called *functional central limit theorem*, which we reproduce below from [30] for convenience. For the theorem to make sense, we first need to construct a sequence[12] $\{S_n^*\}_{n\in\mathbb{N}}$ of random functions on the unit interval $[0,1]$. Let $\{X_k\}_{k\in\mathbb{N}}$ be a sequence of i.i.d random variables with mean 0 and variance 1. Call their partial sums $S_n \coloneqq \sum_{k=1}^{n} X_k$. The sequence $\{S_n\}_{n\in\mathbb{N}}$ now represents a random walk on $\mathbb{N}$. If we embed $\mathbb{N} \hookrightarrow [0,\infty)$, we may construct a function $S : [0,\infty) \to \mathbb{R}$ by interpolating linearly between the points of the random walk:

$$S(t) \coloneqq S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor)(S_{\lfloor t \rfloor + 1} - S_{\lfloor t \rfloor}), \quad t \in [0,\infty)$$

In layman's terms, Donsker's invariance principle states that any random walk with normalized i.i.d. increments may be considered a standard Brownian motion[13] if time and space are infinitely 'shrunk' in the appropriate ratio. This 'shrinking' procedure is reflected in the following sequence of random functions. For any $t \in [0,1]$ and $n \in \mathbb{N}$, define:

$$S_n^*(t) \coloneqq \frac{S(nt)}{\sqrt{n}} \tag{2.12}$$

The argument nt represents the time scaling and the denominator $\sqrt{n}$ represents the space scaling of the embedded random walk S.

Theorem 2.2.1 (Donsker's invariance principle[14])**.** *On the space* $\boldsymbol{C}[0,1]$ *of continuous functions on the unit interval with the metric induced by the sup-norm, the sequence* $\{S_n^* : n \geq 1\}$ *converges in distribution to a standard Brownian motion* $\{B(t) : t \in [0,1]\}$.

[12] The letter S in this section is *unrelated* to our state variable from the rest of this work. We u

[13] This is the same as a standard Wiener process.

[14] Reproduced from [30]

The result from the theorem states that $B(t)$ is defined on $[0, 1]$. This is not a real constraint since Brownian motion possesses a scaling invariance property. For any $a > 0$, $\frac{1}{a}B(a^2t)$ is a standard Brownian motion if and only if $B(t)$ is a standard Brownian motion. This allows us to redefine the time interval of the resulting process.

Donsker's invariance principle is relevant to our problem because the decision variable from the SPRT (equation (2.10)) is exactly a random walk with i.i.d. increments $\{I_k\}_{k\in\mathbb{N}}$. However, these increments are not normalized. Calling $m \coloneqq \mathbb{E}[I_k]$ and $D^2 \coloneqq \mathrm{Var}(I_k)$ for all $k \in \mathbb{N}$, we rewrite the evidence variable as:

$$y_n = D \cdot \sum_{k=1}^{n} \frac{I_k - m}{D} + \sum_{k=1}^{n} m \tag{2.13}$$

Note how we have divided the accumulation process $\{y_n\}_{n\in\mathbb{N}}$ into two random walks. The first sum represents the fluctuating nature of the sensory evidence and gives rise to the $D \cdot dW_t$ term in the DDM (equation (2.11)). The second sum is not stochastic at all and gives rise to the drift part of the DDM.

In equation (2.13), the summands from the first sum in the right-hand side now have zero mean and unit variance. Donsker's invariance principle therefore applies. Nevertheless, a outstanding issue remains, which is that the deterministic term $n \cdot m$ resulting from the second sum in (2.13) does not behave well under the aforementioned space-time scaling procedure. This is not surprising, as it is a well-known fact in stochastic calculus that Brownian motion and deterministic time don't have equivalent differentials. A common rule of thumb is that dW_t is comparable to $\sqrt{dt}$. Hence, Donsker's invariance principle should *not* be invoked for this linear function of n. Rather, a choice has to be made as to what this term becomes in continuous time. If time is scaled in the same way as in (2.12), then $m \cdot n$ becomes $m \cdot n \cdot t$ and any space rescaling of order smaller than n would result in an exploding process (reaching infinity in finite time with probability one). Similarly, a rescaling of order greater than n will make the continuous process vanish with probability one. The only remaining meaningful scaling is one of order $O(n)$. Note, however, that this specification leaves as class of candidate scalings any scalar multiple of n, $\{\alpha \cdot n : \alpha > 0\}$.

Interpretation of the decision variable in the DDM

An important point of interpretation must be stressed regarding the link between the SPRT and the DDM. In the SPRT, the evidence variable y_n *is* the log posterior odds

ratio. In the DDM, the notion of a particular realization of the sensory input $\{o_t\}_{t\geq 0}$ is lost. Rather, the continuous time evidence variable y_t may only be interpreted as having the *distribution* of an ensemble of ideal observers sampling independent observations *from the same environment*. Note also that for Gaussian likelihoods (f_i) the decision variable from the DDM is a scaled version of the log posterior odds ratio, as explained in [10] and [32].

2.2.2 Sampled-time approximation

When going from a discrete time to a continuous time setting, the mathematical models for the environment and observations need to be adapted. From being a discrete time Markov chain $\{S_n\}_{n\in\mathbb{N}}$ with transition matrix $\left(\eta^{ij}\right)_{(i,j)\in\Lambda^2}$, the environment becomes a continuous time Markov chain $\{S_t\}_{t\geq 0}$ with infinitesimal generator $\left(\eta^{ij}\right)_{(i,j)\in\Lambda^2}$. In particular, the following two conditions hold:

- $\eta^{ii} = -\sum_{j\in\Lambda\setminus\{i\}} \eta^{ji}$ for all $i \in \Lambda$
- $\eta^{ij} \in [0,\infty)$ for all $(i,j) \in \Lambda^2$ satisfying $i \neq j$

The correspondence between a discrete time and a continuous time conception of the environment can be made precise via a *sampled-time approximation* scheme, which we now describe.

Let $\{S_t\}_{t\geq 0}$ be the continuous time Markov chain just described. For small $0 < \Delta t \ll 1$, we construct a discrete time chain $\left\{S_n^{\Delta t}\right\}_{n\in\mathbb{N}}$ with transition rates defined as follows[15],

$$p_{\Delta t}^{ij} \coloneqq \begin{cases} \Delta t \cdot \eta^{ij} & \text{if } i \neq j \\ 1 - \Delta t \cdot \sum_{i\in\Lambda\setminus\{j\}} \eta^{ij} & \text{if } i = j \end{cases} \tag{2.14}$$

We then construct the continuous time process $\{S_t^{\text{approx}}\}_{t\geq\Delta t}$ as a sampled-time approximation of $\{S_t\}_{t\geq 0}$ by setting,

$$S_t^{\text{approx}} \coloneqq S_n^{\Delta t}$$

[15]Note that to avoid any confusion regarding the notation for the different transition rates and probabilities involved, in the current derivation, we reserve the notation η^{ij} for the transition rates of the continuous time chain, and $p_{\Delta t}^{ij}$ for the transition probabilities of the discrete time chain.

for every pair $(t, n) \in (0, \infty) \times \mathbb{N}$ satisfying $n\Delta t \leq t < (n+1)\Delta t$. In other words, $\{S_t^{\text{approx}}\}_{t \geq \Delta t}$ is the embedding of $\left\{S_n^{\Delta t}\right\}_{n \in \mathbb{N}}$ in continuous time, with (potential) jumps only occurring at integer multiples of Δt. In-between these times, S_t^{approx} is constant. It is known that sampling $\{S_t\}_{t \geq 0}$ at discrete time steps Δt yields an embedded discrete time chain $\{S_{n\Delta t}\}_{n \in \mathbb{N}}$ whose transition rates only differ from (2.14) by $o(\Delta t)$. In all the continuum limit procedures performed in the remainder of this work, the discrete time decision making algorithms from section 2.1 are applied to the chain $\{S_n^{\Delta t}\}_{n \in \mathbb{N}}$ and embedded in continuous time via $\{S_t^{\text{approx}}\}_{t \geq \Delta t}$. Taking the limit $\Delta t \to 0$ then turns the discrete time recursive equations from section 2.1 into SDEs (or ODEs for pulsatile evidence) that are applicable to $\{S_t\}_{t \geq 0}$.

Remark: For the remainder of the dissertation, as $n\Delta t$ proves cumbersome to read in long derivations, we instantiate the following notation,

$$t_n \coloneqq n\Delta t, \qquad \forall \Delta t > 0, \forall n \in \mathbb{N}$$

The reader must therefore keep in mind that every symbol t_n in this work implicitly involves a choice of Δt. Also, it will be customary for us to both let $\Delta t \to 0$ and $n \to \infty$ under the constraint that $n\Delta t \to t$ for some $t \in [0, T]$. We will always only write $\Delta t \to 0$ instead, trusting that this will not hamper clarity.

2.2.3 Changing 2-state environment with asymmetric rates

In this section, we take the continuum limit of the recursive system (2.8), which is the extension of the SPRT to a 2-state changing environment. Because the first two terms in the right-hand side of equation (2.8) are the same as for the SPRT, their continuum limit will be similar to the DDM just derived, with the exception that drift and diffusion will now be time-dependent. Indeed, in a changing environment, the sequential information I_k gathered at each time step has mean (and potentially variance) that depend on the environmental state S_k. More precisely, consecutive blocks of observations $I_k, I_{k+1}, \ldots, I_{k+\ell}$ are i.i.d only during epochs in which the environment is constant: $S_k = S_{k+1} = \cdots = S_{k+\ell}$. Thus, the use of the DDM derivation applies separately on each such epoch, and the resulting continuous time evidence variable y_t ultimately obeys a SDE of the form,

$$dy_t = m_t \cdot dt + D_t \cdot dW_t + g(y_t) \cdot dt \tag{2.15}$$

where m_t and D_t are stepwise constant between change-points. Just as for the DDM, we usually set the deterministic initial condition $y_0 = 0$. The function g in the

SDE (2.15) above, comes from the last term in the right-hand side of equation (2.8a). We derive its explicit form below.

Because we have established that the term I_k in equation (2.8) gives rise to the drift and diffusion coefficients from equation (2.15), it is sufficient to consider the following difference equation to derive the explicit form of g.

$$\tilde{y}_n - \tilde{y}_{n-1} = \log \frac{\eta^{10} \cdot \exp(-\tilde{y}_{n-1}) + \left(1 - \eta^{01}\right)}{\eta^{01} \cdot \exp(\tilde{y}_{n-1}) + (1 - \eta^{10})}$$

As per section 2.2.2, we embed this equation in continuous time:

$$\tilde{y}_{t_n} - \tilde{y}_{t_{n-1}} = \log \frac{\Delta t \cdot \eta^{10} \cdot \exp(-\tilde{y}_{t_{n-1}}) + \left(1 - \Delta t \cdot \eta^{01}\right)}{\Delta t \cdot \eta^{01} \cdot \exp(\tilde{y}_{t_{n-1}}) + (1 - \Delta t \cdot \eta^{10})}$$

Rearranging and approximating the log terms linearly in time about 1 yields:

$$\begin{aligned}
\tilde{y}_{t_n} - \tilde{y}_{t_{n-1}} &= \log \left[\Delta t \left(\eta^{10} \cdot \exp(-\tilde{y}_{t_{n-1}}) - \eta^{01}\right) + 1\right] \\
&\qquad - \log \left[\Delta t \left(\eta^{01} \cdot \exp(\tilde{y}_{t_{n-1}}) - \eta^{10}\right) + 1\right] \\
&= \Delta t \left(\eta^{10} \cdot \exp(-\tilde{y}_{t_{n-1}}) - \eta^{01}\right) \\
&\qquad - \Delta t \left(\eta^{01} \cdot \exp(\tilde{y}_{t_{n-1}}) - \eta^{10}\right) + o(\Delta t)
\end{aligned} \tag{2.16}$$

Finally, we divide both sides of equation (2.16) by Δt and let $\Delta t \to 0$. Assuming that this noiseless evidence $\tilde{y}_t$ evolves *continuously* in time, we obtain the Ordinary Differential Equation (ODE):

$$\frac{d\tilde{y}_t}{dt} = \eta^{10} \left(\exp(-\tilde{y}_t) + 1\right) - \eta^{01} \left(\exp(\tilde{y}_t) + 1\right) =: g(\tilde{y}_t) \tag{2.17}$$

Note how this last equation gives us the explicit form of g, which may now be plugged back into (2.15). An equivalent derivation as the one presented above can be found in [49].

2.2.4 Changing 2-state environment with symmetric rates

Here, the environment S_t is a continuous time Markov chain with state space $\Lambda = \{0, 1\}$ and symmetric exit rate, $h \in (0, \infty)$, from each state. We proceed in very much the same way as we did to derive the discrete time decision making algorithm for a

symmetric environment (equation (2.9)) from the asymmetric case (equation (2.8)). Setting $h \coloneqq \eta^{10} = \eta^{01}$, the function g from (2.17) becomes,

$$g(y_t) = -h\left(\exp(y_t) - \exp(-y_t)\right) = -2h \cdot \sinh(y_t)$$

and the SDE (2.15) simplifies to:

$$dy_t = m_t dt + D_t dW_t - 2h \cdot \sinh(y_t)dt \tag{2.18}$$

Again, a deterministic initial condition, $y_0 = 0$, is usually enforced. This decision making model was extensively studied in [49], and also used in [17, 23].

2.2.5 Ideal observer models with pulsatile evidence

We derive here the ideal observer models for the Poisson clicks task and the dynamic clicks task presented in the introduction chapter.

Poisson clicks task (static environment)

Let us remind the set-up for the Poisson clicks task, while introducing our notation. An observer receives two independent streams of Poisson clicks during the time interval $[0, T]$. Calling λ_L and λ_R the arrival rates for the left and right streams, respectively, the sets of left and right click times, $\{\ell_i\}_{i\in L}$ and $\{r_i\}_{i\in R}$, satisfy the following conditions.

- $R \coloneqq \{1, \dots, n_R\}$ for some $n_R \sim Poi(\lambda_R \cdot T)$. If $n_R = 0$, then $R \coloneqq \emptyset$.
- $L \coloneqq \{1, \dots, n_L\}$ for some $n_L \sim Poi(\lambda_L \cdot T)$. If $n_L = 0$, then $L \coloneqq \emptyset$.
- The inter-event times of each stream are i.i.d. exponential random variables. More precisely, setting $r_0 = \ell_0 = 0$ and $r_{n_R+1} = \ell_{n_L+1} = T$, we have $r_{i+1} - r_i \sim Exp(1/\lambda_R)$ and $\ell_{j+1} - \ell_j \sim Exp(1/\lambda_L)$ for all $i \in R$ and $j \in L$.

On any given trial, either $(\lambda_L, \lambda_R) = \left(\lambda_{\text{low}}, \lambda_{\text{high}}\right)$ or $(\lambda_L, \lambda_R) = \left(\lambda_{\text{high}}, \lambda_{\text{low}}\right)$, for some constant rates $\lambda_{\text{high}} > \lambda_{\text{low}} > 0$. The task consists in reporting "left" if $(\lambda_L, \lambda_R) = \left(\lambda_{\text{high}}, \lambda_{\text{low}}\right)$, and "right" otherwise, which is equivalent to identifying the stream with highest click rate.

To cast this decision making problem into the SPRT framework, we let the binary random variable $S \in \{0, 1\}$ govern the assignment of each arrival rate to its side as

follows. When $S = 0$, then $(\lambda_L, \lambda_R) = \big(\lambda_{\text{high}}, \lambda_{\text{low}}\big)$, and when $S = 1$, the reverse assignment holds. Furthermore, we discretize time into countably many contiguous bins of width Δt, applying the sampled-time approximation scheme from section 2.2.2 to the observation process $\{o_t\}_{t\geq 0}$ itself[16]. Observation o_{t_n} in time bin n is defined as the pair of click counts for each side during the time window $[n\Delta t,\ (n+1)\Delta t)$. Denoting by $\texttt{1o}, \texttt{o1}, \texttt{oo}$, the events that, respectively, a left click, a right click and no click occurs in the time bin, we define our observation space as $\Xi \coloneqq \{\texttt{1o}, \texttt{o1}, \texttt{oo}\}$. Since the streams are independent and have Poisson statistics, for small $0 < \Delta t \ll 1$, the probability that more than a single click, across both streams, fall within a time bin has probability[17] $o(\Delta t)$. We are therefore justified in our above definition of Ξ.

We now derive the likelihoods of each observation, given the underlying state of the environment. To emphasize the dependence of these functions on the time step size, we write Δt as a superscript. Thus, $f_i^{\Delta t}(o) \coloneqq \Pr\Big(o \;\Big|\; S = i\Big)$, for all $(i, o) \in \Lambda \times \Xi$.

$$\begin{aligned} f_1^{\Delta t}(\texttt{1o}) &= [\lambda_L \Delta t + o(\Delta t)][1 - (\lambda_R \Delta t + o(\Delta t))] \\ &= \lambda_L \Delta t + o(\Delta t) \\ &= \lambda_{\text{low}} \Delta t + o(\Delta t) \end{aligned}$$

$$\begin{aligned} f_1^{\Delta t}(\texttt{o1}) &= [1 - (\lambda_L \Delta t + o(\Delta t))][\lambda_R \Delta t + o(\Delta t)] \\ &= \lambda_R \Delta t + o(\Delta t) \\ &= \lambda_{\text{high}} \Delta t + o(\Delta t) \end{aligned}$$

$$\begin{aligned} f_1^{\Delta t}(\texttt{oo}) &= [1 - \lambda_L \Delta t + o(\Delta t)][1 - \lambda_R \Delta t + o(\Delta t)] \\ &= 1 - (\lambda_{\text{low}} + \lambda_{\text{high}})\Delta t + o(\Delta t) \end{aligned}$$

Similar derivations hold for the $S = 0$ condition. After dropping the $o(\Delta t)$ terms as per section 2.2.2, we obtain the following likelihood functions.

$$f_1^{\Delta t}(\texttt{o1}) = f_0^{\Delta t}(\texttt{1o}) = \lambda_{\text{high}} \Delta t \tag{2.19a}$$

$$f_1^{\Delta t}(\texttt{1o}) = f_0^{\Delta t}(\texttt{o1}) = \lambda_{\text{low}} \Delta t \tag{2.19b}$$

$$f_1^{\Delta t}(\texttt{oo}) = f_0^{\Delta t}(\texttt{oo}) = 1 - (\lambda_{\text{low}} + \lambda_{\text{high}})\Delta t \tag{2.19c}$$

[16] Here, each stream is a Poisson process, and as such, a continuous time Markov chain.

[17] The reader might have noticed the slight ambiguity in our use of the symbol o. We apologize for it and hope that this note will clarify any potential remaining doubt. Whenever this symbol is used on its own, or with a time subscript, as in o and o_{t_n}, it means an observation. Whenever it is used with Δt as an argument, as in $o(\Delta t)$, it represents the Landau "little-oh" notation.

The log likelihood ratio from equation (2.7) becomes (with $n \in \mathbb{N}$),

$$I_{t_n} = \begin{cases} -\kappa \coloneqq \log \dfrac{\lambda_{\text{low}}}{\lambda_{\text{high}}} & \text{if } o_{t_n} = \texttt{10} \\ \kappa \coloneqq \log \dfrac{\lambda_{\text{high}}}{\lambda_{\text{low}}} & \text{if } o_{t_n} = \texttt{01} \\ 0 = \log \dfrac{1-(\lambda_{\text{low}}+\lambda_{\text{high}})\Delta t}{1-(\lambda_{\text{low}}+\lambda_{\text{high}})\Delta t} & \text{if } o_{t_n} = \texttt{00} \end{cases} \tag{2.20}$$

where κ may be interpreted as the magnitude of the evidence gained at the occurrence of each click. At this point, the SPRT equation (2.10) embedded in continuous time takes the form:

$$y_{t_n} = I_{t_n} + y_{t_{n-1}} \qquad \forall n \in \mathbb{N} \tag{2.21a}$$

$$y_0 \coloneqq y_{t_0} \coloneqq \log \frac{\pi(S=1)}{\pi(S=0)} \tag{2.21b}$$

Its continuum limit may finally be taken by subtracting off $y_{t_{n-1}}$, dividing through by Δt, and letting $\Delta t \to 0$, in (2.21a):

$$\frac{dy_t}{dt} = \kappa \left(\sum_{i \in R} \delta\left(t - r_i\right) - \sum_{j \in L} \delta\left(t - \ell_j\right) \right) \tag{2.22a}$$

$$y_0 = \log \frac{\pi(S=1)}{\pi(S=0)} \tag{2.22b}$$

Equation (2.22) is a "jump" ODE for the decision variable of our ideal observer model. The solution to this ODE has jump discontinuities at every click time. Every right click provokes an upward jump of size κ in the decision variable, and every left click generates a downward jump of similar size. In contrast with the decision variable from the DDM, y_t, here, *is exactly* the log posterior odds ratio of the two alternatives $S \in \{0, 1\}$. This was achievable because the evidence arriving from the observation stream is sparse enough: Only clicks carry additional information about the state of the environment. This also illustrates a fundamental difference in the type of noise from the two settings. With continuously arriving evidence, as in the DDM, we assume that the noise occurrs in the observation space. In a sense, this could be considered measurement noise in space. In the case of pulsatile evidence, however, the noise is temporal. Brunton et al. [12] have studied models combining these different types of noise. For instance, a click could be missed or mislocalized

with some probability, and the decision variable itself could be subject to random fluctuations. Although interesting, these models fall outside the scope of our present dissertation.

Dynamic clicks task (dynamic environment)

In the dynamic clicks task, the underlying state S_t undergoes switches during the trial according to the same Markovian dynamics as the ones described in section 2.2.4. Using the same approach as for the Poisson clicks task, we embed equation (2.9) in continuous time.

$$y_{t_n} = I_{t_n} + y_{t_{n-1}} + \log \frac{h \cdot \exp(-y_{t_{n-1}}) + (1-h)}{h \cdot \exp(y_{t_{n-1}}) + (1-h)} \qquad \forall n > 1 \tag{2.23a}$$

$$y_{t_1} = I_{t_1} + \log \frac{\pi(S=1)}{\pi(S=0)} \tag{2.23b}$$

Notice that the first part of this equation is identical to equation (2.21), while the last term is identical to the deterministic part of section 2.2.4. Therefore, the continuum limit ultimately becomes:

$$\frac{dy_t}{dt} = \kappa \left(\sum_{i \in R} \delta\left(t - r_i\right) - \sum_{j \in L} \delta\left(t - \ell_j\right) \right) - 2h \cdot \sinh(y_t) \qquad \forall t \geq 0 \tag{2.24a}$$

$$y_0 \coloneqq \log \frac{\pi(S=1)}{\pi(S=0)} \tag{2.24b}$$

In the continuum limit above, as $\Delta t \to 0$, we have $t_1 \to 0$, $I_{t_1} \to 0$, and $y_{t_1} \to y_0$, with y_0 defined in equation (2.24b). It is also implicitly understood that the transition probability $h \in (0,1)$ in (2.23a), converges to the hazard rate $h \in (0,\infty)$ in (2.24a). Finally, each click stream, considered individually, is an inhomogeneous Poisson process. Therefore, the cardinalities of the indexing sets R and L will not be Poisson random variables anymore. Nevertheless, each set still indexes the click times of the individual streams.

Equations (2.24) describe the evolution of the log posterior odds ratio of an ideal observer during a trial of the dynamic clicks task. An example trajectory for an artificial[18] trial is illustrated in figure 2.1. At click times, the log posterior

[18]By this we mean that we artificially enforced a state transition at 0.5 seconds, in a 1-second long trial, for illustration purposes.

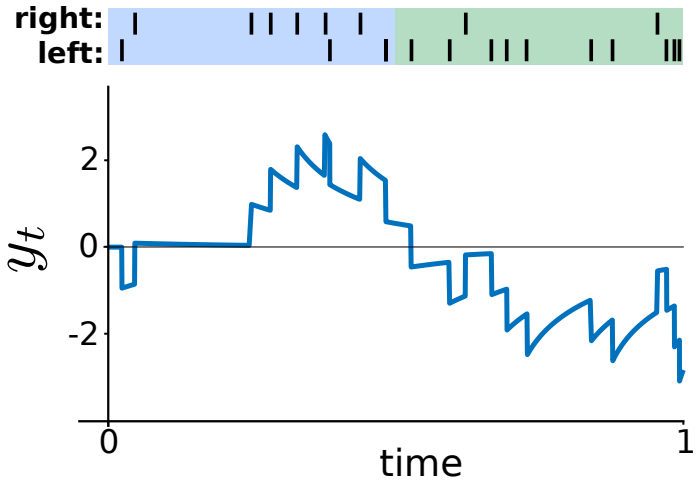

Figure 2.1: **Single trajectory of the log likelihood ratio y_t, during an artificial trial of the dynamic clicks task.** The decision variable (blue trace) evolves according to equations (2.24). Discontinuous jumps occur at click times in the direction corresponding to the side of the click, and a nonlinear decay to 0 characterizes the periods in-between clicks. Poisson clicks for the right and left streams (black ticks), and environmental state (background color), are depicted above the graph.

odds ratio jumps discontinuously in a direction prescribed by the click's side. The constant magnitude of the jumps is controlled by the log ratio of the click rates κ (equation (2.20)). In-between clicks, the evidence decays nonlinearly to 0. Just as for equation (2.22), the decision variable from equation (2.24) is a deterministic function of the click times and model parameters. We wish to emphasize the similarity in the leak term $-2h \cdot \sinh(y_t)$, with our previous ideal observer model for continuously arriving evidence (equation (2.18)).

2.3 Summary

Important concepts and techniques from normative decision making theory were introduced. In discrete time 2AFC tasks, classical models describe the evolution in time of a decision variable, the accrued *evidence*, which represents the log posterior odds ratio of the two choices, given the history of observations. A continuum limit of such models yields SDEs in continuous time. For static environments, such models

are often referred to as perfect integrators.

In dynamic environments undergoing Markovian dynamics, a fixed discounting function may be applied, on-line, to the running evidence. This effectively implements a "forgetting" effect, by which old observations bear less on the present decision than recent ones. In an ideal observer model, the sharpness of the evidence decay is tuned to the environment's volatility – the hazard rate – and the magnitude of the evidence. That is, both a highly volatile environment or a large amount of evidence will dampen integration. This behavior enables the ideal observer model to be more sensitive to change-points than a traditional perfect integrator.

The continuous time decision variable from our SDE models *is not*, per se, a log posterior odds ratio. The reason is that the stimulus itself becomes a diffusion process. Conditioning on a single realization of it is meaningless (a single realization has probability 0 for a stochastic process with continuous state space). The correct way of thinking about these continuous time normative models is in terms of their distribution. For a fixed realization of the hidden environmental chain, the distribution of the dynamic variable from the SDE represents that of an ensemble of log posterior odds ratios, coming from independent ideal observers. Continuous time normative models, therefore, remain useful for investigating statistical properties of the decision process, such as the accuracy of the response or the expected decision time.

Finally, we showed how the normative framework above could be adapted to the case of pulsatile evidence. This "sparseness" of the evidence, compared to the diffusion process mentioned above, allows the continuum limit to faithfully represent the log posterior odds ratio.

Chapter 3

Decision making with pulsatile evidence

In the past few years, the dynamic clicks task was developed in the lab of Carlos Brody [34], as an extension of the Poisson clicks task [12] to changing environments. These tasks have the particularity, together with a few other ones [33, 35, 18], to deliver pulsatile evidence to the subject, with exponentially distributed inter-pulse intervals. Such stimuli depart from the random dots tradition [44, 11, 46, 42, 23], in which evidence arrives continuously in time. Although ideal observer models for pulsatile evidence integration appear in the literature, their mathematical analysis is still incomplete.

In section 2.2.5 we derived the ideal observer model for the dynamic clicks task. In the present chapter, we analyze this model and a linear version of it, from the perspective of their choice accuracy. Choice accuracy – percentage correct – is indeed a common dependent variable in behavioral experiments, and as such, constitutes a natural first step in the exploration of the models.

3.1 SNR in the Poisson clicks task

The Signal-to-Noise Ratio (SNR) mathematically captures the difficulty level of a decision making task. It is, as such, a useful tool in the analysis of ideal observer models [49]. Depending on the field of study and the problem at hand, several definitions have been used for this quantity. It is commonly defined as the ratio of

signal power to noise power [20], the ratio of signal variance to noise variance [15], and the ratio of the mean stimulus to its standard deviation [29]. Following Veliz-Cuba and colleagues [49], we use this latter definition.

In the Poisson clicks task, the stimulus is composed of two Poisson streams, $\left(\{\ell_i\}_{i\in L}, \{r_i\}_{i\in R}\right)$. Taking the mean of such object may be done in several ways. To define the SNR we first project the stimulus onto a one-dimensional space, by defining it as the difference in click counts between the right and left streams. The intuition behind this projection is that, with no prior bias on the most probable side for the high-rate stream, the optimal strategy is to select the side with highest click count. With the notation introduced in section 2.2.5, this means that $\hat{o}_T \coloneqq n_R - n_L$. This projected stimulus is the difference of two Poisson random variables, and as such, it is distributed according to the Skellam distribution [47]. In particular, the ratio of its mean to standard deviation is given by,

$$\text{SNR} = \frac{T\left(\lambda_{\text{high}} - \lambda_{\text{low}}\right)}{\sqrt{T\left(\lambda_{\text{high}} + \lambda_{\text{low}}\right)}} = \sqrt{T}\frac{\lambda_{\text{high}} - \lambda_{\text{low}}}{\sqrt{\lambda_{\text{high}} + \lambda_{\text{low}}}} =: \sqrt{T}\cdot\mathcal{S}, \tag{3.1}$$

where we have defined the discriminability index $\mathcal{S}$ [19, 43]. As we will see in the next few sections, both parameters T and $\mathcal{S}$ shape the accuracy on the dynamic clicks task, provided an adequate time rescaling is applied.

3.2 Time rescaling in the dynamic clicks task

The ideal observer model for the dynamic clicks task (equation (2.24)) involves four parameters: the two click rates $\lambda_{\text{low}}, \lambda_{\text{high}}$, the hazard rate h, and the trial duration T. In order to ease the analysis of this model, we rescale time in a way that the hazard rate becomes 1. For simplicity of exposition, assume that the original time unit[1] in equation (2.24) is the second. A hazard rate of h Hz means that, on average, the environment undergoes h change-points per second. Let's call our time variable τ in our new time unit, and t in the old one (so t is expressed in seconds). Because we want, on average, one change-point per units of τ, the scaling must take the form,

$$\tau = \frac{t}{h}$$

Thus, one unit of τ equals $1/h$ seconds, and one second equals h units of τ. Table 3.1 summarizes the transformation that all our parameters undergo under this time

[1] In reality, any time unit works, it doesn't have to be seconds.

rescaling. We add the subscript τ to the parameter symbol when it is expressed in units of τ. Finally, we point out that another perspective may be taken on this time

parameter name	**param. symbol/value (time in sec.)**	**param. symbol/value (time in units of τ)**
hazard rate	h	$h^\tau = 1$
low click rate	λ_{low}	$\lambda_{\text{low}}^\tau = \lambda_{\text{low}}/h$
high click rate	λ_{high}	$\lambda_{\text{high}}^\tau = \lambda_{\text{high}}/h$
trial duration	T	$T^\tau = h \cdot T$
left click times	ℓ_i	$\ell_i^\tau = h \cdot \ell_i$
right click times	r_i	$r_i^\tau = h \cdot r_i$
discriminability index	$\mathcal{S}$	$\mathcal{S}^\tau = \mathcal{S}/\sqrt{h}$
SNR	SNR	$\text{SNR}^\tau = \text{SNR}$
jump size	κ	$\kappa^\tau = \kappa$

Table 3.1: **Parameters from the ideal observer model for the dynamic clicks task, expressed in two different time units.** The time variable t may represent seconds, while the time variable τ is such that one unit of τ equals $1/h$ seconds. Note that SNR and κ are not affected by this transformation.

rescaling procedure. Because T seconds correspond to $h \cdot T$ units of τ, and h is itself in units of $1/t$, one may consider that τ is a *unitless*, i.e., *non-dimensionalized*, unit of time. This is a common procedure in Physics[2] to reduce the dimensionality of a problem. The non-dimensionalized model now reads,

$$\frac{dy_\tau}{d\tau} = \kappa \left(\sum_{i \in R} \delta\left(\tau - r_i^\tau\right) - \sum_{j \in L} \delta\left(\tau - \ell_j^\tau\right) \right) - 2\sinh(y_\tau) \tag{3.2}$$

with similar initial condition as (2.24b).

[2]A similar time rescaling appears in equation (3.5) from [49], for instance.

3.3 Effective parameters in the dynamic clicks task

Accuracy is an evident dependent variable to measure in a decision making experiment. By accuracy, we mean the fraction of trials (between 0 and 1, or in percentages) on which the decision maker makes the correct choice. It is an overall measure of performance and reflects the difficulty level of a task. Experimenters heavily rely on this variable to calibrate their experiment to each individual. In sensory neuroscience, it is customary to tune task parameters so that each individual reaches intermediate accuracy, in order to best reveal the cognitive or sensory processes at stake.

The time rescaling procedure from section 3.2 has reduced the number of core parameters to three: the rescaled click rates $\lambda^\tau_{\text{low}}, \lambda^\tau_{\text{high}}$, and rescaled trial duration T^τ. The important question that remains is: *what combination of parameters determines the accuracy?* Numerical exploration revealed that, for high enough values of $\lambda^\tau_{\text{high}}$ and $\lambda^\tau_{\text{low}}$, keeping the two parameters T^τ and $\mathcal{S}^\tau$ constant was sufficient to keep accuracy constant. In other words, in the non-dimensionalized model (3.2), and for high enough click rates, only *two* parameters determine the model's accuracy. Figure 3.1 illustrates this result.

3.4 Best linear model and sensitivity analysis

Although equation (3.2) describes the accumulation process of an ideal observer, it is possible that a non-optimal observer will perform with similar accuracy[3] in practice. So, we now compare the optimal model to a simpler model. Following [49] and [34], we examine a linear model of evidence accumulation in which the nonlinear *sinh*-term from equation (3.2) is replaced by a linear term, proportional to the accumulated evidence:

$$\frac{dy_\tau}{d\tau} = \kappa \left[\sum_{i \in R} \delta(\tau - r_i^\tau) - \sum_{j \in L} \delta(\tau - \ell_j^\tau) \right] - \gamma y_\tau \tag{3.3a}$$

$$y_0 = \log \frac{\pi(S=1)}{\pi(S=0)} \tag{3.3b}$$

[3]Recall that accuracy is the percentage of correct trials.

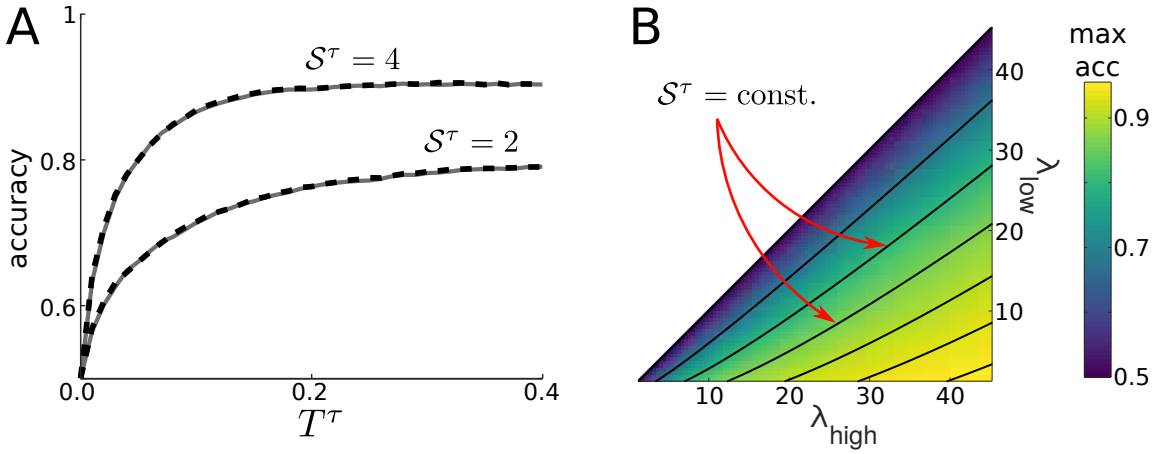

Figure 3.1: **Two effective parameters determine accuracy. A:** Choice accuracy of the ideal observer model from equation (2.24) as a function of rescaled interrogation time T^τ, for two distinct values of $\mathcal{S}^\tau$ (see table 3.1). Two distinct pairs of click rates were used in simulations ($\lambda^\tau_{\text{low}} = 30$ and 60 Hz and $\lambda^\tau_{\text{high}}$ picked to match the appropriate $\mathcal{S}^\tau$ level), yielding the overlaying solid and dashed lines. Keeping T^τ and $\mathcal{S}^\tau$ fixed yields a match in performance for both pairs of click rates. As time evolves during the trial, performance saturates to a maximal value less than 1. **B:** Numerically computed maximal accuracy (color level) of the ideal observer model is constant along the level curves (black lines) of $\mathcal{S}^\tau$ (seen as a function of $(\lambda^\tau_{\text{low}}, \lambda^\tau_{\text{high}})$). We only display the first octant of the $\lambda^\tau_{\text{high}}\lambda^\tau_{\text{low}}$-plane because of the requirement, $0 < \lambda^\tau_{\text{low}} < \lambda^\tau_{\text{high}}$.

If $\gamma = 2$, this new system becomes the piecewise[4] linearization of (3.2) at the origin. However, both [49] and [34] have shown that this linearization is *not* the one that yields maximal accuracy on the decision task at hand[5]. Instead, it is possible to tune the discounting parameter in (3.3) to an accuracy-maximizing value γ^*. The resulting accuracy *closely matches the performance of the optimal model*. This phenomenon may be appreciated[6] in figure 3.2A. Importantly, figure 3.2B shows how γ^* increases with $\mathcal{S}^\tau$. This is not the case for the nonlinear model, in which the optimal discounting parameter h^τ is constantly 1.

[4]Meaning, in-between click times.

[5]We are considering here the dynamic clicks task under the interrogation protocol, just as [34]. Veliz-Cuba and colleagues [49] study the linearization of the SDE (2.18) in section 2.2.4.

[6]We assume that T^τ is large enough so that accuracy has saturated, as in figure 3.1C.

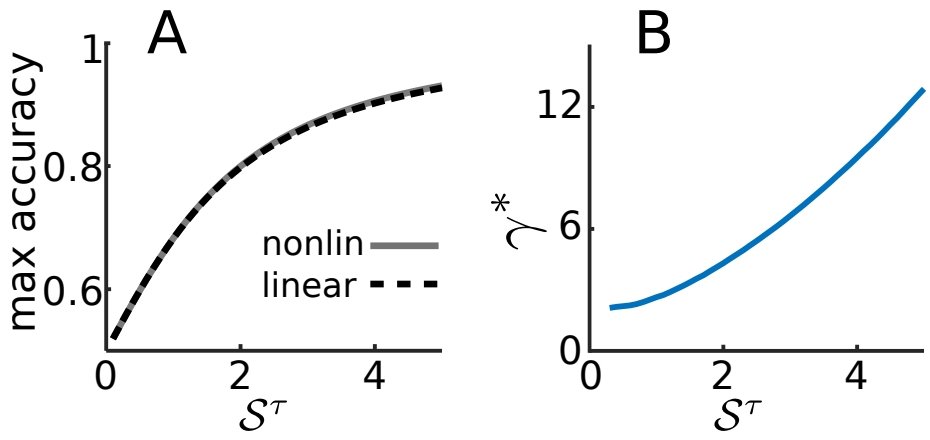

Figure 3.2: **Accuracy of tuned linear model matches the one from the ideal observer. A:** Strong agreement in performance between the nonlinear and linear model (with best γ^*) for a wide range of $\mathcal{S}^\tau$ values. **B:** Best linear discounting rate γ^* for equation (3.3) as a function of $\mathcal{S}^\tau$ (defined in table 3.1).

Remark: In this section, the expressions "nonlinear model", "optimal model", and "ideal observer model", all refer to the same system from equation (3.2). We consider its discounting parameter to be h^τ. The alternative system (3.3) will always be called the "linear model", and its discounting parameter is γ.

We now perturb the discounting parameter of both models around its maximal-accuracy value. Our goal is to analyze the effects of such perturbation on choice accuracy. We apply the perturbation relative to the magnitude of the optimal parameter value. For the nonlinear model, the optimal value is $h^\tau = 1$, and for the linear model, it is $\gamma = \gamma^*$, which must be found numerically. So, a relative perturbation of $c \in \mathbb{R}$, means the following,

$$\tilde{\gamma}(c) \coloneqq \gamma^* + c\gamma^* \tag{3.4a}$$

$$\tilde{h}(c) \coloneqq 1 + c \tag{3.4b}$$

where $\tilde{\gamma}$ and $\tilde{h}$ are the perturbed parameters. We find that the linear model is more sensitive to the perturbation than the nonlinear model (figure 3.3). We quantify the sensitivity of both models by computing the curvature of their accuracy functions for a range of $\mathcal{S}^\tau$ values (figure 3.3B). The accuracy of the linear model has higher curvature at the optimal parameter value γ^*, indicating higher sensitivity.

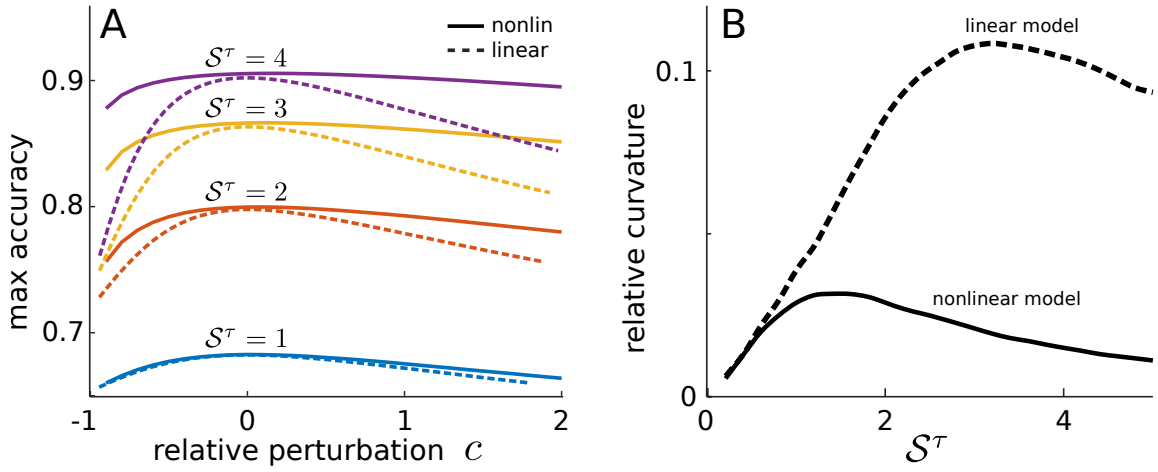

Figure 3.3: **Model's sensitivity to discounting parameter perturbation. A:** Performance profile around the optimal discounting parameter for the nonlinear (solid) and linear (dashed) models, presented at several $\mathcal{S}^\tau$-levels (see annotations on curves, or color coding). The linear model is systematically more sensitive to parameter perturbation than its nonlinear counterpart. The relative perturbation c is defined in equations (3.4). **B:** Curvature (absolute value of the second derivative) of the curves from panel A, evaluated at their peak, as a function of $\mathcal{S}^\tau$. The curvature, and hence the sensitivity, of the linear model clearly separates from the curvature of the nonlinear model, for moderate to high $\mathcal{S}^\tau$-values.

Our analysis not only reveals the sensitivity of the linear model, but also the relative insensitivity of both models at extreme values of $\mathcal{S}^\tau$. This result can be understood intuitively. If $\mathcal{S}^\tau$ is small, the observer will not be able to receive enough evidence to make an informed decision. At high $\mathcal{S}^\tau$, the observer receives strong evidence from a single click, and does not need to integrate or discount old evidence. In either case, accuracy depends weakly on the assumed discounting parameter, and the performances of the two models are close (bottom and top curves of figure 3.3A). We also note that the insensitivity of the nonlinear model to changes in its discounting parameter suggests that this is a more robust model: An observer who does not learn the hazard rate h^τ exactly can still perform well. The linear model requires finer tuning to achieve best performance.

This points towards a key trade-off in using mathematical models of evidence accumulation (or models in general). Optimal models may be more robust, but more difficult to implement. At the same time, due to their robustness, it may not be necessary to tune optimal models for nearly optimal performance. On the

other hand linear models are less robust, but might be easier to implement. In the next section, we will explore these differences between models properties further, by studying how quickly the linear and normative model's discounting parameters can be fitted to choice data.

3.5 Model fitting

Fitting models to experimental data is a practice of paramount importance in contemporary science. It is a central part of the scientific method through which a coherent understanding of the natural world is built. When a model is well fit to data, it gains predictive power over the studied phenomenon. From there, the model becomes useful in several ways. In the first place, it functions as a source of inspiration for the scientist. From it, one might infer fundamental principles at work in the topic of study and generate new research hypotheses. In the second place, a well-fit model may lead to the design of new technology.

Nevertheless, the methods to use to fit models to data and judge of the goodness-of-fit are far from creating unanimity among the scientific community. Not only is there debate on how to fit a model to data, but also on how to compare fits from different models.

Our definition of accuracy is straightforward. It is the percentage of trials on which an observer makes the correct choice. One may distinguish the *empirical* from the *theoretical* accuracy. The empirical accuracy of an observer may be computed on a specific (finite) set of trials. On the other hand, the theoretical accuracy may be thought of as an expectation. For fixed model and task parameters, such accuracy corresponds exactly to the probability of the observer to be correct on any given trial. When the statistics of the task and the decision making model are known, it is possible, in simple cases, to compute the theoretical accuracy analytically. For instance, when the decision making model accrues sensory information into a decision variable such as the log posterior odds ratio, the accuracy is the probability that the decision variable ends the trial on the correct side of the decision boundary. This is why, when the evolution of the decision variable is described by a stochastic differential equation, solving the Fokker-Planck equation is a common method to yield an analytical formula for the accuracy.

In most cases, however, an analytical expression is not within reach, and one turns to numerical simulations to estimate it. Assuming that producing synthetic trials and running the inference algorithm on these can be done, estimating accuracy

is then as simple as recording the choices made by the algorithm on a large number of i.i.d. trials, and dividing the number of correct choices by the total number of trials.

In this section, we investigate the fit to choice data of the decision making models for pulsatile evidence previously presented. Since the linear and nonlinear models yield comparable accuracy on the dynamic clicks task, we wonder whether they differ in some respect when it comes to fitting them to choice data[7]. Thus, our goal is to present the fitting problem in the dynamic clicks task for these two models, and to provide some numerical results regarding the speed of convergence of the fitted discounting parameter.

3.5.1 The models

In this work, we fit a single parameter from each model to choice data; the leak rate γ for the linear model, and the hazard rate h for the nonlinear one. The equations for these two models, together with their noisy version, are presented below. For simplicity, we drop the notation of time rescaling. However, we keep the hazard rate of the stimulus to 1 in all our simulated data, so that all the interpretation may still be done with time rescaling in mind. All initial conditions in the ODEs to come are the same: $y_0 = 0$ (with probability one in the noisy versions).

The deterministic linear model presented in section 3.4 is:

$$\frac{dy_t}{dt} = \kappa \left[\sum_{i \in R} \delta(t - r_i) - \sum_{j \in L} \delta(t - \ell_j) \right] - \gamma \cdot y_t \tag{3.5a}$$

The deterministic nonlinear model from section 2.2.5 is:

$$\frac{dy_t}{dt} = \kappa \left(\sum_{i \in R} \delta\left(t - r_i\right) - \sum_{j \in L} \delta\left(t - \ell_j\right) \right) - 2h \cdot \sinh(y_t) \tag{3.5b}$$

There are several ways in which we can introduce noise into our models. Common examples are additive or multiplicative noise on either of the clicks evidence or the decision variable, mislocalization[8] noise on clicks, and a lapse rate. However, every noise type comes with its own additional free parameters regarding the fitting procedure, and it is the modeler's task to find the appropriate trade-off between a model's

[7] Here, all the data used are produced synthetically via numerical simulation.

[8] In appendix B, we treat the case of mislocalization (or mis-attribution) noise on clicks.

expressive power and its complexity. In this work, we make the choice of progressing incrementally from the deterministic case. We believe that the issues at hand with fitting our decision making models can already be revealed with simple models. We consider the case of multiplicative Gaussian noise applied to clicks evidence, as it is one of the noise sources that appears in the literature [12, 34]. The stochastic linear model is,

$$\frac{dy_t}{dt} = \sum_{i \in R} \xi_i \cdot \delta(t - r_i) - \sum_{j \in L} \xi_j \cdot \delta(t - \ell_j) - \gamma \cdot y, \tag{3.6}$$

where ξ_i, ξ_j are all i.i.d. Gaussian random variables with mean κ and known standard deviation σ. With noise similarly defined as for the linear model, the stochastic nonlinear model is:

$$\frac{dy_t}{dt} = \sum_{i \in R} \xi_i \delta\left(t - r_i\right) - \sum_{j \in L} \xi_j \delta\left(t - \ell_j\right) - 2h \cdot \sinh(y_t) \tag{3.7}$$

For all models, at fixed interrogation time T, a decision is made according to the following rule:

- if $y_T > 0$, the right stream ($S = 1$) is chosen,
- if $y_T < 0$, the left stream ($S = 0$) is chosen.

In section 2.1.2, we argued that the optimal strategy for the $y_T = 0$ case was to choose one of the two alternatives with probability 1/2. Here, however, we remark that this case has a null probability of occurrence on trials containing at least one click. Furthermore, trials containing no click at all cannot provide any information about the evidence discounting parameter, because these trials are simply void of evidence. Hence, although these trials are useful in experiments – to assess noise and response bias for instance, or just as control trials – we dismiss them altogether from our data. As a result, we assume for the remainder of this chapter that $\Pr(y_T = 0) = 0$ in all models, and that the decision outputted is always a deterministic function of the decision variable's end-value.

3.5.2 Bayesian parameter estimation

All our fitting procedures may be interpreted under the same overarching theory of Bayesian parameter estimation. Let θ denote the free parameter that we want to

fit; i.e., θ represents γ for the linear model and h for the nonlinear one. During a simulated experiment, we collect[9] data from N i.i.d. trials,

$$\mathfrak{D} := \left\{ (\mathcal{T}_k, d_k) : 1 \leq k \leq N \right\} \tag{3.8}$$

where $\mathcal{T}_k := \left(\{\ell_i\}_{i \in L}, \{r_i\}_{i \in R} \right)$ is the clicks stimulus from trial k, and $d_k \in \{0, 1\}$ is the decision (or choice) datum for this trial. Our goal is to compute or estimate the posterior distribution $\Pr\left(\theta \mid \mathfrak{D}\right)$, which by Bayes' rule is proportional to the product of the likelihood of the data with the prior over the parameter[10]:

$$\Pr\left(\theta \mid \mathfrak{D}\right) \propto \Pr\left(\mathfrak{D} \mid \theta\right) \pi(\theta) \tag{3.9}$$

Our method focuses on exploiting the likelihood function $\Pr\left(\mathfrak{D} \mid \theta\right)$. We have,

$$\begin{aligned} \Pr\left(\mathfrak{D} \mid \theta\right) &= \Pr\left(\mathcal{T}_{1:N}, d_{1:N} \mid \theta\right) \\ &= \Pr\left(d_{1:N} \mid \mathcal{T}_{1:N}, \theta\right) \Pr\left(\mathcal{T}_{1:N} \mid \theta\right) \\ &= \Pr\left(d_{1:N} \mid \mathcal{T}_{1:N}, \theta\right) \Pr\left(\mathcal{T}_{1:N}\right), \end{aligned}$$

where the last step comes from the fact that the clicks trains are independent of the discounting parameter θ used by the decision making model. From there, we remark that the choice data are conditionally independent on the clicks stimulus and the discounting parameter. Thus,

$$\Pr\left(d_{1:N} \mid \mathcal{T}_{1:N}, \theta\right) = \prod_{k=1}^{N} \Pr\left(d_k \mid \mathcal{T}_k, \theta\right)$$

The above derivations enable us to rewrite equation (3.9) as:

$$\Pr\left(\theta \mid \mathfrak{D}\right) \propto \pi(\theta) \prod_{k=1}^{N} \Pr\left(d_k \mid \mathcal{T}_k, \theta\right) \tag{3.10}$$

In the rest of the section, we will always use uniform priors for θ, over a finite interval $[0, a]$. In this context, the problem of computing the posterior distribution of θ

[9] Meaning, "we generate". We use the verb "collect" in an attempt to metaphorize an experimental procedure.

[10] Since all the other task and model parameters are assumed known and fixed, we may omit them from the equations.

reduces to assessing the likelihoods of the decision data on each trial, $\Pr\left(d_k \mid \mathcal{T}_k, \theta\right)$ $(1 \leq k \leq N)$, for a range of θ-values spanning the interval $[0, a]$. Finally, note that for numerical stability reasons, our algorithms will sum log-likelihood values, as opposed to multiplying probability values. Relegating the θ-independent prior into a normalization constant C, equation (3.10) becomes, in the log-domain:

$$\log \Pr\left(\theta \mid \mathfrak{D}\right) = C + \sum_{k=1}^{N} \log \Pr\left(d_k \mid \mathcal{T}_k, \theta\right), \quad \theta \in [0, a] \tag{3.11}$$

3.5.3 Error in parameter space

Working with synthetic data, we have perfect knowledge of the model used to produce the data. Because we are fitting two models of decision making, we decided to produce the data with these same models. More specifically, let $m_f, m_d \in \{\mathrm{L}, \mathrm{NL}\}$ represent the model that we fit and the model used to produce the decision data, respectively – L for "linear" and NL for "nonlinear". In the present work, we study the four possible combinations of pairs (m_f, m_d).

To judge of the quality of our fit, we need to define an error function. We define this function in parameter space as the relative mean posterior squared error, averaged across experiments.

$$\mathrm{err}\left(p_1, \ldots, p_M, \theta_{\mathrm{true}}\right) = \frac{1}{\theta_{\mathrm{true}}^2 M} \sum_{i=1}^{M} \int_0^{\infty} p_i(\theta)(\theta - \theta_{\mathrm{true}})^2 d\theta \tag{3.12}$$

In equation (3.12), M is an arbitrary number of experiments, and p_i denotes the posterior density for our parameter θ, obtained under fitting procedure[11] i. The parameter θ_{true} may be defined in two separate ways. When $m_f = m_d$, it is simply the parameter that was used to produce the choice data. However, when $m_f \neq m_d$, a bijection $h \to \gamma$ between hazard rate and leak rate must be used. Setting $h = 1$ in all our simulations, we may use the numerically produced curve from figure 3.2B, representing a bijection $h \to \gamma^*$. When fitting the linear model to data produced with the nonlinear model, we define θ_{true} as the best leak rate γ^* for $h = 1$ and the particular value of $\mathcal{S}$ chosen. Similarly, when fitting the nonlinear model to data produced with the linear model, we define $\theta_{\mathrm{true}} = 1$ while producing the choice data with γ^*. Finally, the factor $1/\theta_{\mathrm{true}}^2$ in the equation explains why we call this error

[11] In this work, we use interchangeably the terms "experiment" and "fitting procedure".

"relative". It ensures that the true parameter be 1, thereby allowing us to compare the errors in both parameter spaces, hazard rate space and leak rate space.

Having defined an error function, we may now study its convergence properties as a function of block size. The block size meta-parameter N is not apparent in equation (3.12). It represents the number of trials used in each fitting procedure to yield the posteriors p_i $(1 \leq i \leq M)$. Presumably, the error should be a decreasing function of N, and we would like to know how the speed of convergence to 0 depends on model pair (m_f, m_d).

3.5.4 Deterministic model fits

For the deterministic models (3.5), the evidence variable y_t is a deterministic function of the initial condition y_0, the click reliability κ, time t, the click times $\mathcal{T}$, and the discounting parameter θ. The decision rule being itself deterministic (recall our prior discussion of the $y_T = 0$ case), the decision output of these models is also deterministic. That is, $\Pr\left(d_k \mid \mathcal{T}_k, \theta\right) \in \{0, 1\}$ $(1 \leq k \leq N)$. Thus, the likelihood of a decision given the click times and the discounting parameter, seen as a function of θ, is a linear combination of indicator functions. For the linear model, we will prove that the support of such likelihood is a union of disjoint intervals.

Fitting the deterministic linear model

The solution to the initial value problem (3.5a) at interrogation time is:

$$y_T = \kappa \sum_{i \in R} \exp\left(-\gamma\left(T - r_i\right)\right) - \kappa \sum_{j \in L} \exp\left(-\gamma\left(T - \ell_j\right)\right) \tag{3.13}$$

As explained above, given the stimulus and choice data from a single trial $(\mathcal{T}, d)$, our goal is to find the set of γ's that is compatible with d. We define the following

decision function:

$$\begin{aligned} dec(\gamma; y_0, T, \mathcal{T}) &= \text{sign}(y_T) \\ &= \text{sign}\left(\kappa \cdot \sum_{i \in R} \exp\left(-\gamma \cdot (T - r_i)\right) - \kappa \cdot \sum_{j \in L} \exp\left(-\gamma \cdot \left(T - \ell_j\right)\right)\right) \\ &= \text{sign}\left(\kappa \cdot \exp\left(-\gamma \cdot T\right)\left[\sum_{i \in R} \exp\left(\gamma \cdot r_i\right) - \sum_{j \in L} \exp\left(\gamma \cdot \ell_j\right)\right]\right) \\ &= \text{sign}\left[\sum_{i \in R} \exp\left(\gamma \cdot r_i\right) - \sum_{j \in L} \exp\left(\gamma \cdot \ell_j\right)\right] \end{aligned}$$

Thus,

$$\gamma \text{ is compatible } \iff dec(\gamma; y_0, T, \mathcal{T}) = d$$

The function

$$\gamma \mapsto \sum_{i \in R} \exp\left(\gamma \cdot r_i\right) - \sum_{j \in L} \exp\left(\gamma \cdot \ell_j\right)$$

is continuous on $[0, \infty)$. Its sign may, therefore, be reached through a characterization of its roots. The general answer is not trivial. If we perform the change of variables $x := \exp(\gamma)$, the problem is equivalent to finding the roots of the following exponential polynomial [41],

$$P(x) = \sum_{i \in R} x^{r_i} - \sum_{j \in L} x^{\ell_j}$$

In general, such polynomial may admit an arbitrary number of roots, which would yield an arbitrary number of disjoint admissible intervals for γ. In practice, however, we impose a closed interval of the form $[0, a]$ as support for the prior $\pi(\gamma)$. Thus, the ranges of a that we explore are small enough for us to directly assess $dec(\gamma; y_0, T, \mathcal{T})$ via equation (3.13), for an array of γ-values evenly spaced across the support. Algorithm 1 below is the one we used to fit the deterministic linear model.

Fitting the deterministic nonlinear model

There is no equivalent of equation (3.13) for the nonlinear model (3.5b). However, equation (3.5b) may still be solved explicitly in-between clicks. If t_1, t_2 denote the

Algorithm 1 Deterministic fit

1: Generate/load $\mathfrak{D} = (\mathcal{T}_k, d_k)_{k=1}^N$ ▷ clicks and decision data
2: Initialize $L \in \mathbb{N}$, $a \in (0, \infty)$
3: *samples_list* $\leftarrow (0, a/L, 2a/L, \ldots, a)$ ▷ candidate θ's span $[0, a]$
4: *compatible* $\leftarrow (True, \ldots, True)$
5: **for** $\mathcal{T}_k \in \mathfrak{D}$ **do**
6: **for** $\theta \in$ *samples_list* **do**
7: Compute $dec(\theta)$
8: $compatible(\theta) \leftarrow dec(\theta) == d_k$
9: **end for**
10: *backup_list* $\leftarrow$ *samples_list*
11: Trim incompatible samples from *samples_list*
12: **if** *samples_list* $== \emptyset$ **then**
13: *samples_list* $\leftarrow$ *backup_list*
14: Refine *samples_list*
15: **end if**
16: **end for**
17: Compute *admissible_intervals* from *samples_list* ▷ disjoint union of intervals
18: **return** *admissible_intervals*

click times of two consecutive clicks, we have:

$$y_{t_2} = 2\text{acoth}\left(\exp\left(2ht_2\right)\coth\frac{y_{t_1}}{2}\right) \tag{3.14}$$

This, or any numerical integration method for the ODE (3.5b), may be used iteratively between clicks to realize line 7 in algorithm 1.

3.5.5 Stochastic model fits

Fitting the linear model

Our noisy linear model of evidence integration was introduced in equation (3.6). In a similar fashion as equation (3.13), the evidence variable at time T is given by,

$$y_T = \sum_{i \in R} \xi_i \cdot \exp\left(-\gamma \cdot (T - r_i)\right) - \sum_{j \in L} \xi_j \cdot \exp\left(-\gamma \cdot \left(T - \ell_j\right)\right) \tag{3.15}$$

We notice that, conditioned on $\{y_0, T, \mathcal{T}\}$, the final evidence value $y(T)$ may be seen as a sum of $n+m$ i.i.d. Gaussian random variables (with $n := |R|$ and $m := |L|$). It is therefore, itself, a Gaussian random variable with mean and variance:

$$\mathbb{E}\left[y_T \middle| y_0, T, \mathcal{T}\right] = \kappa \left[\sum_{i \in R} \exp\left(-\gamma \cdot (T - r_i)\right) - \sum_{j \in L} \exp\left(-\gamma \cdot \left(T - \ell_j\right)\right)\right] \quad (3.16)$$

$$\mathrm{Var}\left(y_T \middle| y_0, T, \mathcal{T}\right) = \sigma^2 \left[\sum_{i \in R} \exp\left(-2\gamma \cdot (T - r_i)\right) + \sum_{j \in L} \exp\left(-2\gamma \cdot \left(T - \ell_j\right)\right)\right] \quad (3.17)$$

The integral of its density over the appropriate domain will, eventually, yield the likelihood of a particular decision given the data from a single trial. That is, for $d = 1$, one integrates the Gaussian density with mean and variance given by equations (3.16) and (3.17), over the positive reals, and for $d = 0$, over the negative reals.

Fitting the nonlinear model

Our estimation of the choice likelihoods for the stochastic version of the nonlinear model requires more work. Rather than attempting an analytical route, we turn to Monte-Carlo sampling. For a given clicks stimulus, and a given realization of the Gaussian multiplicative noise, equation (3.14) allows us to assess the end-value of the model's evidence y_T. We call each such value a *particle*. To estimate the likelihood of a decision, we generate N_{part} particles over independent realizations of the noise, for fixed click times and θ-value. Figure 3.4 shows how the likelihood, estimated as the fraction of particles landing on the side of the y-axis that is compatible with the decision datum, stabilizes as a function of the number of particles. Based on these results, we choose $N_{\text{part}} = 800$ as a trade-off between precision on the estimated likelihood and computational time. As a second validation of our sampling method, we tested it on the linear model and compared it against the analytical solution from section 3.5.5. Figure 3.5 represents a superposition of the Gaussian density from section 3.5.5 on the histogram of landing values of y_T.

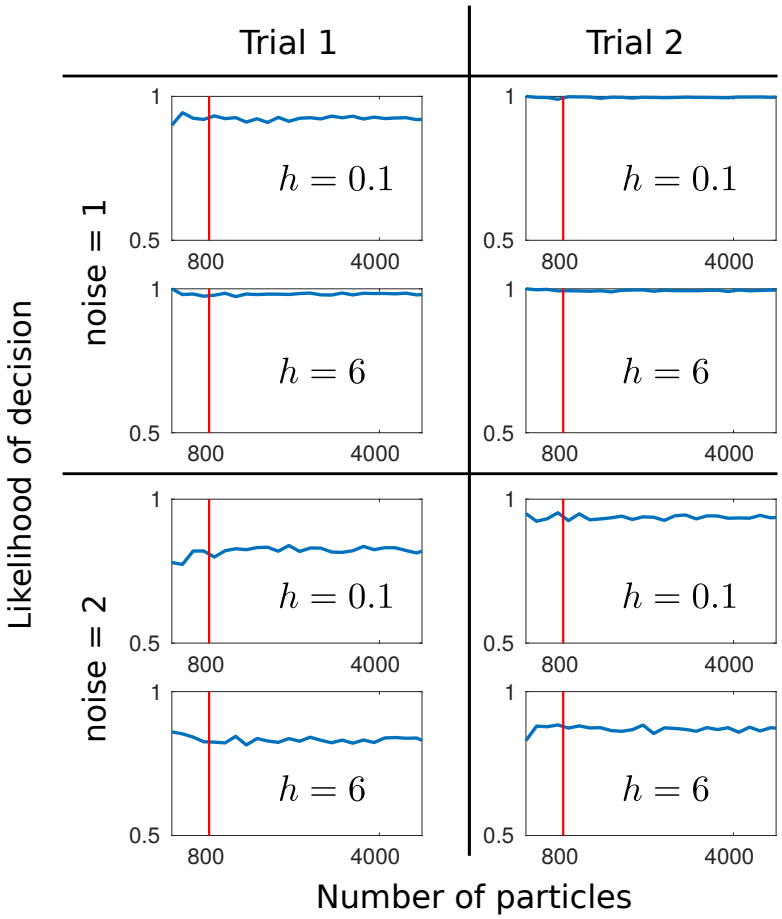

Figure 3.4: **Stabilization of the likelihood estimate yielded by our Monte-Carlo sampling algorithm, as the number of particles increases.** The estimated likelihood (blue curves) is given by the fraction of particles that lands on the correct side of the decision boundary ($y \equiv 0$), in evidence space. The algorithm is applied to the stochastic nonlinear model (3.7), for two distinct trials (columns), two noise values (rows), and two hazard rates (sub-divisions of rows). We used independent particles for distinct points on the x-axis. The red vertical lines indicate the number of particles that we settled on for the other simulations (figures 3.5 and 3.6).

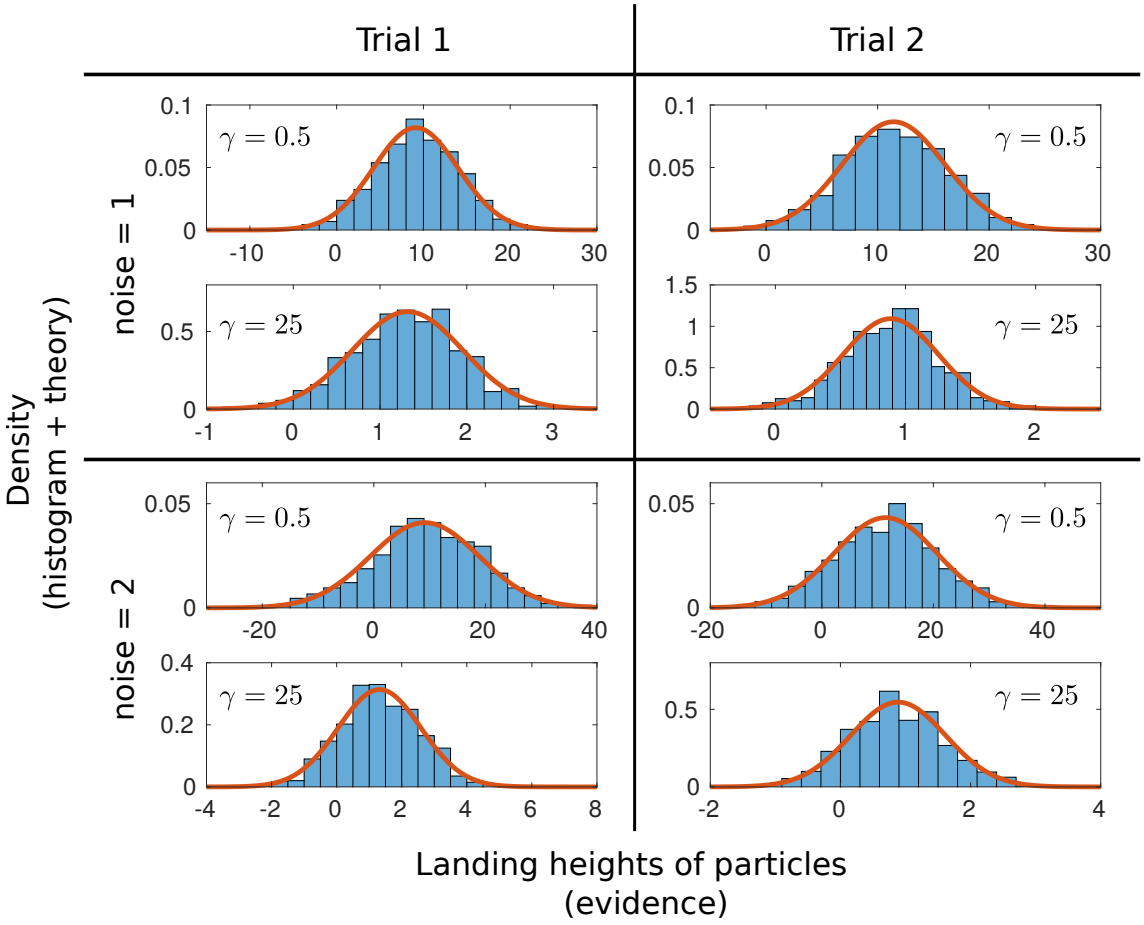

Figure 3.5: **Validation of our Monte-Carlo sampling algorithm.** For two random trials (columns), two noise values (rows), and two discounting parameter values (sub-divisions of rows), we superimpose the theoretical Gaussian density from section 3.5.5 on the histogram of landing heights of 800 particles. Each particle samples the multiplicative noise applied to clicks height independently, and evolves according to equation (3.6). We observe agreement between the theoretical densities and the histograms. Note: we used our sampling algorithm on the linear model, here, only to be able to compare its end-results to available theoretical densities.

3.5.6 Results

The results of our simulated fitting procedures across the four models are summarized in figure 3.6. The relative error defined in section 3.5.3 is plotted against block size

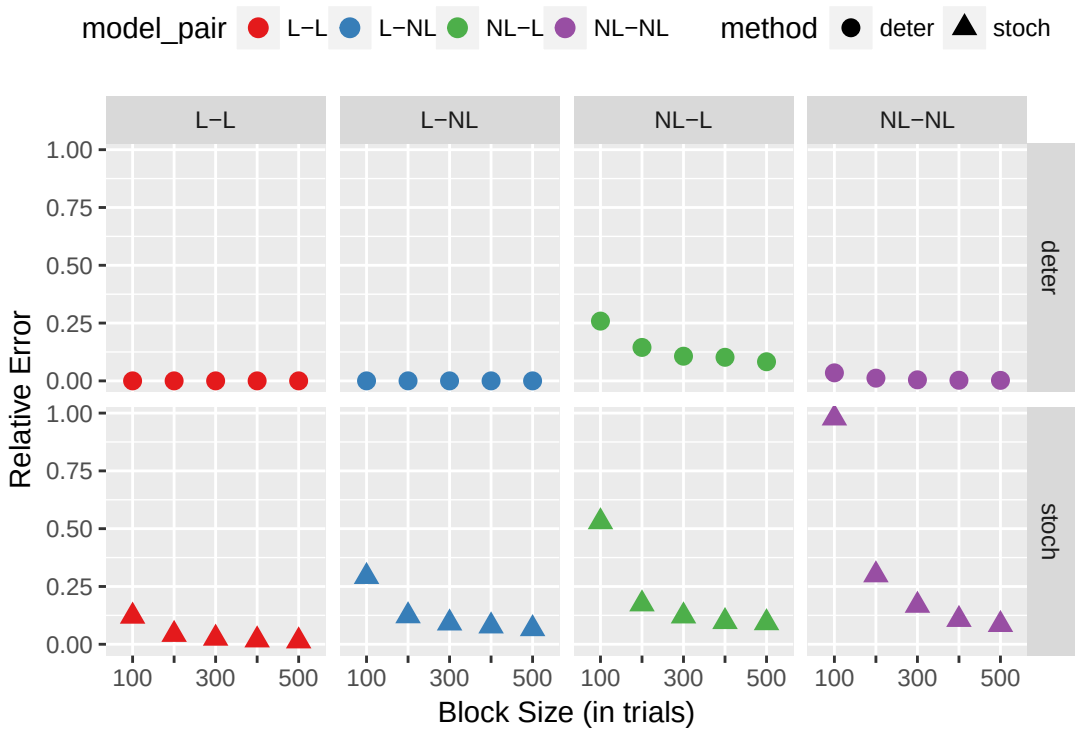

Figure 3.6: **Relative error of fitting procedures as a function of block size.** Each column (color-coded) corresponds to a specific pair m_f-m_d, where L and NL stand for "linear" and "nonlinear" respectively. Each row (coded by symbol shape) refers to the version of the models used to, both, produce the data and perform the fit. For instance, the panel containing the blue triangles means that we fitted the stochastic linear model to choice data synthetically produced by the stochastic nonlinear model. Error is decreasing in every panel. It decreases faster for deterministic models than for stochastic models, and also faster for the fits of the linear model than for the fits of the nonlinear model.

N. Each point is computed across $M = 500$ independent blocks.

We see that the relative error in parameter space decreases consistently quicker when fitting the linear model, compared to fitting the nonlinear model. Note that

such result does *not* necessarily imply that the linear fits better predict choice data than the nonlinear fits. Out-of-sample validations would be required for this, and we suggest this path for further investigations.

3.6 Summary

In this chapter, we performed an in-depth analysis of several key aspects of the ideal observer model for the dynamic clicks task. After suggesting a definition for the SNR of the stimulus, we showed that it involved two important parameters: the trial duration and the discriminability index of the two Poisson click rates. In a changing environment, the analysis of the ideal observer model benefits from a time nondimensionalization. With time units rescaled by the hazard rate, an *effective hazard rate* of 1 reduces the number of parameters of the model. Importantly, we found by numerical exploration that the rescaled trial duration and discriminability index are the only two parameters that determine choice accuracy in high click rate regimes.

Previous studies of normative decision making in changing environments have noted that an appropriately tuned linear model can reach near-optimal accuracy in a wide range of experimentally relevant parameter regimes [23, 49, 34]. We investigated the question of how such linear model differs from the ideal observer model, in terms of sensitivity to parameter perturbation. Using the drop from optimal accuracy as our measure of sensitivity, we found that the linear model is *more sensitive* than the nonlinear model. That is, for equal perturbation[12] in the discounting parameter, the drop in accuracy of the linear model is more pronounced than for the nonlinear model.

Finally, we asked whether the two decision models differ when it comes to fitting their discounting parameter to choice data. Our synthetic fitting procedures suggest that the linear model is easier to fit than the nonlinear model. By "easier to fit", we mean that it takes fewer data trials for the linear model than for the nonlinear model, to recover their discounting parameter with a given tolerance. These results hold not only for the ODE models from chapter 2, but also when multiplicative Gaussian noise is added to each click's evidence.

[12] Measured in percentage of the best parameter value.

Chapter 4

Learning the change rates in discrete time

In the previous chapters, we saw how ideal observer models possessing full knowledge of the generative model[1] are able to use all the information contained in the stimulus in order to maximize choice accuracy. Since experimental subjects are usually rewarded on the basis of this latter quantity, it would be ideal for the subjects to use such models. However, assuming full knowledge of the generative model is too strong an assumption to be realistic [39]. At least during the training period of an experiment, for example, the subject has little knowledge of the latent temporal statistics of the stimulus. It is thus crucial to understand how the generative model, or at least parameters of it, may be *learned*. In the present chapter, we show how an ideal observer can learn the change rates of a discrete time environment. This finds direct applications for the triangles task [23] presented in chapter 1, and is closely related to the learning model from [22].

[1] The generative model is the set of statistical rules by which the stimulus is generated.

4.1 Learning a single hazard rate

4.1.1 In a 2-state environment

Joint inference on state and hazard rate

The mathematical strategy to perform a sequential 2AFC task optimally while learning the hazard rate is to compute the *joint posterior* over state and hazard rate: $\Pr\left(S_n, h \mid \mathcal{O}_n\right)$. We will derive below a recursive formula (equation (4.4)) for computing such posterior.

Our derivation does not depend on the domains of the hazard rate and observations. Thus, we only make the following minimal assumptions:

- $h \in H$ and $o \in \Xi$, for some $(H, \Xi) \subseteq [0,1] \times \mathbb{R}$. We are not imposing that H nor Ξ be uncountable, nor infinite.
- The noise distributions $\{f_i\}_{i \in \Lambda}$ are non- vanishing on Ξ.
- For all $n \in \mathbb{N}$ and all $i \in \Lambda$, $\Pr(S_n = i) > 0$.
- Our ideal observer model uses a non-vanishing prior joint density over states and hazard rates, $\pi : \Lambda \times H \to (0, \infty)$.
- The model further considers state and hazard rate to be independent, so that the joint prior factors into the product of its marginals[2]: $\pi(S, h) = \pi(S) \cdot \pi(h)$.

We may now start our derivation of the recursive formula for the joint posterior over state and hazard rate. We start by the initial conditions of the iterative process. After one observation o_1 has been gathered, Bayes' rule gives us:

$$\Pr\left(S_1, h \mid \mathcal{O}_1\right) = \frac{1}{\Pr(\mathcal{O}_1)} f_{S_1}(o_1) \cdot \pi(S_1) \cdot \pi(h) \tag{4.1}$$

For $n > 1$, $h \in H, S_n \in \Lambda$ and a stream $\mathcal{O}_n \in \Xi^n$ of n observations, the law of total probability allows us to marginalize over the previous state value:

$$\Pr\left(S_n, h \mid \mathcal{O}_n\right) = \sum_{i \in \Lambda} \Pr\left(S_n, S_{n-1} = i, h \mid \mathcal{O}_n\right)$$

[2]We are aware of our abuse of notation in using π to denote both the joint distribution and its marginals, but the context will always be sufficiently clear to avoid confusion in the next pages.

Then, Bayes' rule applied to the summand in the right-hand side of the previous equation yields:

$$\Pr\left(S_n, h \mid \mathcal{O}_n\right) = \sum_{i\in\Lambda} \frac{\Pr\left(\mathcal{O}_n \mid S_n, S_{n-1} = i, h\right) \Pr\left(S_n, S_{n-1} = i, h\right)}{\Pr\left(\mathcal{O}_n\right)} \tag{4.2}$$

At this point, we encounter again two facts that will simplify the above summand, just as it was the case in our previous derivation of equation (2.2a). The first fact is that, by assumption, the observations are conditionally independent on the states, and they are also independent from the hazard rate. Hence,

$$\Pr\left(\mathcal{O}_n \mid S_n, S_{n-1} = i, h\right) = f_{S_n}(o_n)\Pr\left(\mathcal{O}_{n-1} \mid S_{n-1} = i, h\right)$$

The second fact is that,

$$\Pr\left(S_n, S_{n-1} = i, h\right) = \Pr\left(S_n \mid S_{n-1} = i, h\right) \Pr\left(S_{n-1} = i, h\right)$$

So, noticing the convenient grouping of terms,

$$\Pr\left(\mathcal{O}_{n-1} \mid S_{n-1} = i, h\right) \Pr\left(S_{n-1} = i, h\right) = \Pr\left(\mathcal{O}_{n-1}\right) \Pr\left(S_{n-1} = i, h \mid \mathcal{O}_{n-1}\right)$$

we may rewrite equation (4.2) as:

$$\begin{aligned}\Pr\left(S_n, h \mid \mathcal{O}_n\right) = \frac{\Pr\left(\mathcal{O}_{n-1}\right)}{\Pr\left(\mathcal{O}_n\right)} f_{S_n}(o_n) \times \\ \sum_{i\in\Lambda} \Pr\left(S_n \mid S_{n-1} = i, h\right) \Pr\left(S_{n-1} = i, h \mid \mathcal{O}_{n-1}\right)\end{aligned} \tag{4.3}$$

Our final step consists in replacing $\Pr\left(S_n \mid S_{n-1} = i, h\right)$ by the very definition of the hazard rate:

$$\begin{aligned}\Pr\left(S_n, h \mid \mathcal{O}_n\right) = \frac{\Pr\left(\mathcal{O}_{n-1}\right)}{\Pr\left(\mathcal{O}_n\right)} f_{S_n}(o_n) \Big[(1-h) \cdot \Pr\left(S_{n-1} = S_n, h \mid \mathcal{O}_{n-1}\right) + \\ h \cdot \Pr\left(S_{n-1} = 1 - S_n, h \mid \mathcal{O}_{n-1}\right)\Big]\end{aligned} \tag{4.4}$$

Remark: In going from equation (4.3) to equation (4.4), we used the fact that $\Lambda = \{0, 1\}$, which implies that $S_{n-1} \neq S_n \Leftrightarrow S_{n-1} = 1 - S_n$.

Equation (4.4), together with its initial condition (4.1), express recursively the posterior over state and hazard rate at time n as a function of the most recent observation o_n and the joint posterior at time $n-1$. An important feature of these equations is that the value of h is the same on both sides of the equality sign. Despite this decoupling in the hazard rate variable, the joint posterior meets the following implicit condition at every time step $n \in \mathbb{N}$:

$$\sum_{i \in \Lambda} \int_H \Pr\left(S_n = i, h \mid \mathcal{O}_n\right) dh = \int_H \sum_{i \in \Lambda} \Pr\left(S_n = i, h \mid \mathcal{O}_n\right) dh = 1 \tag{4.5}$$

Thus, in our current set-up, the ideal observer performs joint on-line inference on state and hazard rate by evolving recursively the family of equations (4.4), indexed by $(S_n, h) \in \Lambda \times H$, with initial conditions prescribed by (4.1).

Joint inference on state and change-point count

An alternative approach to on-line learning of the hazard rate is to carry a probability distribution (belief) over change-point count. A change-point is a transition in the hidden state. Whenever, $S_{n-1} \neq S_n$ $(n > 1)$, we say that a change-point occurs at time step n. This approach was exposed in detail in another context by [51]. We adapt it here to our problem.

Let a_n be the number of change-points, and $b_n = n - 1 - a_n$ the count of non-change-points between times 1 and n $(n = 1, 2, \ldots)$ (see figure 4.1B). The process $\{a_n\}_{n \geq 1}$ is a pure birth process with birth rate h. The observer assumes no changes prior to the start of observation, $\Pr(a_1 = 0) = 1$, and must make at least two observations, o_1 and o_2, to detect a change.

To develop an iterative equation for the joint conditional probability density, $\Pr\left(S_n, a_n \mid \mathcal{O}_n\right)$, given the n observations $\mathcal{O}_n$, we begin by marginalizing over these quantities at the time of the previous observation, $n-1$, for $n > 1$. Bayes' rule and the law of total probability first yield:

$$\Pr\left(S_n, a_n \mid \mathcal{O}_n\right) = \frac{1}{\Pr\left(\mathcal{O}_n\right)} \sum_{S_{n-1} \in \{0,1\}} \sum_{a_{n-1}=0}^{n-2} \Pr\left(\mathcal{O}_n \mid S_n, S_{n-1}, a_n, a_{n-1}\right) \times \Pr\left(S_n, S_{n-1}, a_n, a_{n-1}\right)$$

[3]This section represents a slight modification of content appearing in [38].

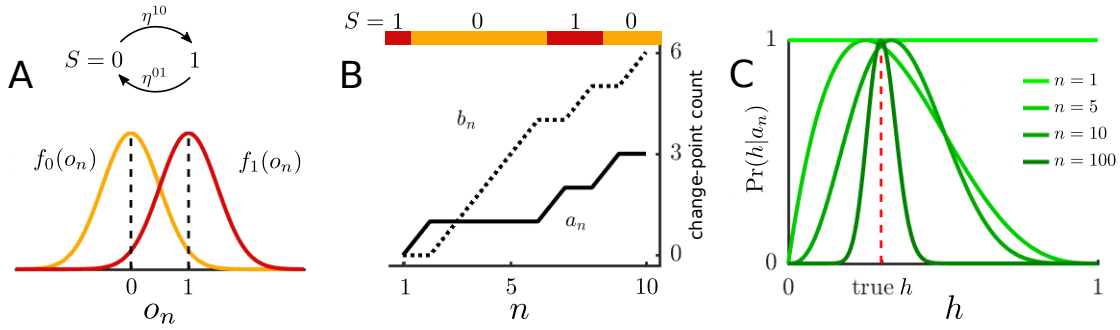

Figure 4.1: **Online inference of the change rate in a dynamic environment (image adapted from [38]). A:** The environment S alternates between states 0 and 1 with transition probabilities η^{01}, η^{10}. We analyze the symmetric case ($h := \eta^{10} = \eta^{01}$) in section 4.1.1 and the asymmetric case $\left(\eta^{01} \neq \eta^{10}\right)$ in section 4.2. The state of the environment determines the sampling distribution of the observations, $f_i(o_n) := \Pr\left(o_n \mid S = i\right)$, which we represent as a Gaussian density. **B:** A sample path of the environment (color bar) together with the first ten values of the actual change-point count (a_n) and non-change-point count (b_n). **C:** Evolution of the conditional probabilities, $\Pr\left(h \mid a_n\right)$ (given by Beta distributions), corresponding to the change-point count from panel B, until $n = 100$. The dashed red line indicates the value of h in the simulation. The densities are scaled so that each equals 1 at the mode.

Using the conditional independence of observations,

$$\Pr\left(\mathcal{O}_n \mid S_n, S_{n-1}, a_n, a_{n-1}\right) = \Pr\left(o_n \mid S_n\right) \Pr\left(\mathcal{O}_{n-1} \mid S_{n-1}, a_{n-1}\right)$$

we find that,

$$\Pr\left(S_n, a_n \mid \mathcal{O}_n\right) = \frac{1}{\Pr\left(\mathcal{O}_n\right)} \sum_{S_{n-1} \in \{0,1\}} \sum_{a_{n-1}=0}^{n-2} \Pr\left(o_n \mid S_n\right) \Pr\left(\mathcal{O}_{n-1} \mid S_{n-1}, a_{n-1}\right) \times \Pr\left(S_n, S_{n-1}, a_n, a_{n-1}\right)$$

Furthermore, we can use the definition of conditional probability to write,

$$\Pr\left(S_n, S_{n-1}, a_n, a_{n-1}\right) = \Pr\left(S_n, a_n \mid S_{n-1}, a_{n-1}\right) \Pr\left(S_{n-1}, a_{n-1}\right)$$

and Bayes' rule also implies,

$$\Pr\left(\mathcal{O}_{n-1} \mid S_{n-1}, a_{n-1}\right) \Pr\left(S_{n-1}, a_{n-1}\right) = \mathrm{P}_{n-1}(S_{n-1}, a_{n-1}) \Pr\left(\mathcal{O}_{n-1}\right)$$

Hence,

$$\Pr\left(S_n, a_n \mid \mathcal{O}_n\right) = \frac{\Pr\left(\mathcal{O}_{n-1}\right)}{\Pr\left(\mathcal{O}_n\right)} \Pr\left(o_n \mid S_n\right) \sum_{S_{n-1}\in\{0,1\}} \sum_{a_{n-1}=0}^{n-2} \Pr\left(S_n, a_n \mid S_{n-1}, a_{n-1}\right) \times \Pr\left(S_{n-1}, a_{n-1} \mid \mathcal{O}_{n-1}\right) \tag{4.6}$$

With two choices we have the following relationships for all $n > 1$:

$$S_n = S_{n-1} \Leftrightarrow a_n = a_{n-1} \quad \text{and} \quad S_n \neq S_{n-1} \Leftrightarrow a_n = a_{n-1} + 1 \tag{4.7}$$

The term $\Pr\left(S_n, a_n \mid S_{n-1}, a_{n-1}\right)$ in equation (4.6) is therefore nonzero only if either, $S_{n-1} = S_n$, and $a_{n-1} = a_n$, or $S_{n-1} \neq S_n$ and $a_{n-1} = a_n - 1$. As illustrated in figure 4.2, if the system is in the joint state (S_{n-1}, a_{n-1}) at time step $n-1$, then at step n it can either, transition to,

$$(S_n \neq S_{n-1}, a_n = a_{n-1} + 1)$$

or remain at,

$$(S_n = S_{n-1}, a_n = a_{n-1})$$

This observation is central to the message-passing algorithm described in [4, 51], with probability mass flowing from lower to higher values of a according to a pure birth process (see figure 4.2). We can thus simplify equation (4.6), leaving only two terms in the double sum. To ease notation, let us define the following notation:

$$\Pr_n(i, a) \coloneqq \Pr\left(S_n = i, a_n = a \mid \mathcal{O}_n\right)$$

Thus, we have for $n > 1$:

$$\begin{aligned}\Pr_n(i, a) = \frac{\Pr\left(\mathcal{O}_{n-1}\right)}{\Pr\left(\mathcal{O}_n\right)} f_i(o_n) \times \\ \Big[\Pr\left(S_n = i, a_n = a \mid S_{n-1} = i, a_{n-1} = a\right) \cdot \Pr_{n-1}(i, a) + \\ \Pr\left(S_n = i, a_n = a \mid S_{n-1} = 1 - i, a_{n-1} = a - 1\right) \cdot \Pr_{n-1}(1 - i, a - 1)\Big]\end{aligned} \tag{4.8}$$

We must also specify *initial conditions* at time 1, and *boundary values* when $a \in \{0, n-1\}$ for these equations. At the first time step, we have $\Pr(a_1 = 0) = 1$. Therefore,

$$\Pr_1(i, 0) = \frac{1}{\Pr(o_1)} f_i(o_1) \pi(S = i) \tag{4.9}$$

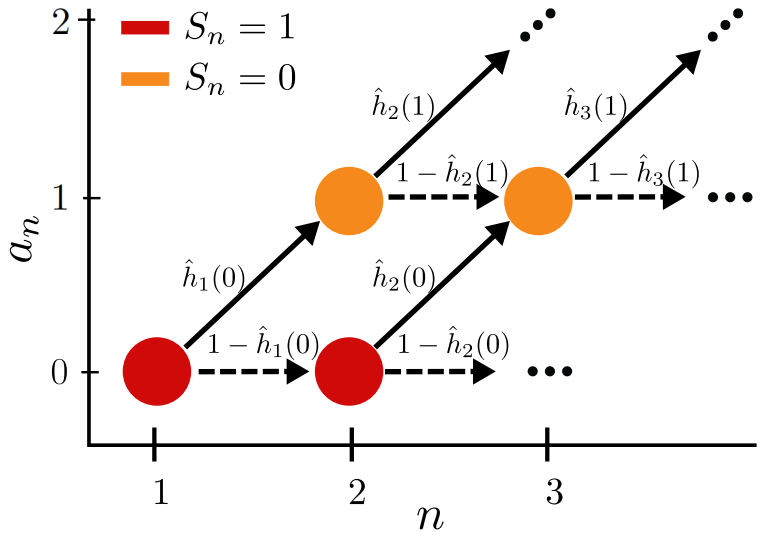

Figure 4.2: **Graph underlying the transport of joint posterior probability mass across time (image adapted from [38]).** The joint posterior over state and change-point count, $\Pr_n(i, a) \coloneqq \Pr(S_n = i, a_n = a \mid \mathcal{O}_n)$, is propagated along a directed graph according to equation (4.19). A colored node in the graph represents a particular point (S_n, a_n) from the support of the joint posterior. The reader should imagine some probability mass associated to each node in a column, during a specific time of the on-line inference process. At the next time step, the mass is transferred across the edges with the appropriate annotated weights (see equations (4.17) and (4.18) for the definition of these weights). For clarity, only paths corresponding to the initial condition $(S_1, a_1) = (1, 0)$ are shown.

and $\Pr_1(i, a) = 0$ for $a \neq 0$. Here $\pi(S = i)$ is the prior over the two choices, which we typically take to be uniform so $\pi(S = 0) = \pi(S = 1)$. The probability $\Pr(o_1)$ is unknown to the observer. However, similar to the ratio $\frac{\Pr(\mathcal{O}_{n-1})}{\Pr(\mathcal{O}_n)}$ in equation (4.8), $\Pr(o_1)$ acts as a normalization constant and does not appear in the posterior odds ratio, R_n (see equation (4.20) below). Finally, at all future times $n > 1$, we have separate equations at the boundaries,

$$\Pr_n(i, 0) = \frac{\Pr(\mathcal{O}_{n-1})}{\Pr(\mathcal{O}_n)} f_i(o_n) \Pr\Big(S_n = i, a_n = 0 \mid S_{n-1} = i, a_{n-1} = 0\Big) \Pr_{n-1}(i, 0) \tag{4.10}$$

and,

$$\Pr_n(i, n-1) = \frac{\Pr(\mathcal{O}_{n-1})}{\Pr(\mathcal{O}_n)} f_i(o_n) \Pr\left(S_n = i, a_n = n-1 \mid S_{n-1} = 1-i, a_{n-1} = n-2\right) \times \Pr_{n-1}(1-i, n-2) \tag{4.11}$$

We next compute $\Pr\left(S_n, a_n \mid S_{n-1}, a_{n-1}\right)$ in equation (4.6), with $n > 1$, by marginalizing over all possible transition rates $h \in [0, 1]$:

$$\Pr\left(S_n, a_n \mid S_{n-1}, a_{n-1}\right) = \int_0^1 \Pr\left(S_n, a_n \mid h, S_{n-1}, a_{n-1}\right) \Pr\left(h \mid S_{n-1}, a_{n-1}\right) dh \tag{4.12}$$

Note that $\Pr\left(h \mid S_{n-1}, a_{n-1}\right) = \Pr\left(h \mid a_{n-1}\right)$, so we need the distribution of h, given a_{n-1} change-points, for all $n > 1$. We assume that prior to any change-point observations — that is at time $n = 1$ — the rates follow a Beta distribution with hyper-parameters $a_0, b_0 > 0$ (see also sections 3.1 and 3.2 in [51]),

$$\pi(h) = \tilde{\beta}(h; a_0, b_0) \coloneqq \frac{h^{a_0-1}(1-h)^{b_0-1}}{B(a_0, b_0)}$$

where $\tilde{\beta}$ denotes the probability density of the associated Beta distribution, and $B(x, y) \coloneqq \int_0^1 s^{x-1}(1-s)^{y-1} ds$ is the Beta function. For any $n > 1$, the random variable $a_n|h$ follows a Binomial distribution with parameters $(n-1, h)$, for which the Beta distribution is a conjugate prior. The posterior over the change rate when the change-point count is known at time $n > 1$ is therefore:

$$h|a_n \sim Beta(a_0 + a_n, b_0 + b_n) \tag{4.13}$$

For simplicity, we assume that prior to any observations, the probability over the transition rates is uniform, $\pi(h) = 1$ for all $h \in [0, 1]$, and therefore $a_0 = b_0 = 1$ (see figure 4.1C).

We now return to equation (4.12) and use the definition of the transition rate, h, (see figure 4.1) to find:

$$\Pr\left(S_n, a_n \mid h, S_{n-1}, a_{n-1}\right) = \begin{cases} 1-h & S_n = S_{n-1} \;\&\; a_n = a_{n-1} \\ h & S_n \neq S_{n-1} \;\&\; a_n = a_{n-1} + 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.14}$$

equation (4.12) can therefore be rewritten using two integrals, depending on the values of (S_n, a_n) and (S_{n-1}, a_{n-1}),

$$\Pr\Big(S_n = i, a_n = a \;\Big|\; S_{n-1} = i, a_{n-1} = a\Big) = \int_0^1 (1-h)\tilde{\beta}(h; a_{n-1}+1, b_{n-1}+1)dh \tag{4.15}$$

and similarly for $\Pr\Big(S_n = i, a_n = a \;\Big|\; S_{n-1} = 1-i, a_{n-1} = a-1\Big)$. The mean of the Beta distribution, for $n > 1$, can be expressed in terms of its two parameters:

$$\hat{h}_{n-1}(a_{n-1}) \coloneqq \mathbb{E}\left[h|a_{n-1}\right] = \frac{a_{n-1}+1}{a_{n-1}+b_{n-1}+2} \tag{4.16}$$

We denote this expected value by $\hat{h}_{n-1}(a_{n-1})$ as it represents a point estimate of the change rate h at time $n-1$ when the change-point count is a_{n-1}, $n > 1$. Since $a_{n-1} + b_{n-1} = n-2$, we have:

$$\hat{h}_{n-1}(a_{n-1}) = \frac{a_{n-1}+1}{n} \tag{4.17}$$

The expected transition rate, $\hat{h}_{n-1}(a_{n-1})$, is thus determined by the ratio between the previous change-point count and the number of time steps, n. Leaving a_0 and b_0 as parameters in the prior gives $\hat{h}_{n-1}(a_{n-1}) = (a_{n-1}+a_0)/(n-2+a_0+b_0)$. Using the definition in equation (4.17), it follows from equation (4.15) that:

$$\Pr\Big(S_n = i, a_n = a \;\Big|\; S_{n-1} = i, a_{n-1} = a\Big) = 1 - \hat{h}_{n-1}(a) \tag{4.18a}$$

$$\Pr\Big(S_n = i, a_n = a \;\Big|\; S_{n-1} = 1-i, a_{n-1} = a-1\Big) = \hat{h}_{n-1}(a-1) \tag{4.18b}$$

Equations (4.18), which are illustrated in figure 4.2, can in turn be substituted into equation (4.8) to yield, for all $n > 1$:

$$\begin{aligned}\Pr_n(i,a) = &\frac{\Pr(\mathcal{O}_{n-1})}{\Pr(\mathcal{O}_n)} f_i(o_n)\Big[\Big(1-\hat{h}_{n-1}(a)\Big)\cdot \Pr_{n-1}(i,a) \\ &+\hat{h}_{n-1}(a-1)\cdot \Pr_{n-1}(1-i,a-1)\Big]\end{aligned} \tag{4.19}$$

The initial conditions and boundary equations for this recursive probability update have already been described in equations (4.9–4.11). Equation (4.19) is the equivalent of equation (3) in [4], and equation (3.7) in [51]. However, here the observer does not need to estimate the length of the interval since the last change-point.

The observer can compute the posterior odds ratio by marginalizing over the change-point count:

$$R_n := \frac{\Pr\left(S_n = 1 \mid \mathcal{O}_n\right)}{\Pr\left(S_n = 0 \mid \mathcal{O}_n\right)} = \frac{\sum_{a=0}^{n-1} \Pr_n(1, a)}{\sum_{a=0}^{n-1} \Pr_n(0, a)} \tag{4.20}$$

Here $\log(R_n) = y_n > 0$ implies that $S_n = 1$ is more likely than $S_n = 0$. Note that $\Pr(\mathcal{O}_{n-1}) / \Pr(\mathcal{O}_n)$ and $1/\Pr(o_1)$ need not be known to the observer to obtain the most likely choice.

A posterior distribution of the transition rate h can also be derived from equation (4.19) by marginalizing over (S_n, a_n),

$$\Pr\left(h \mid \mathcal{O}_n\right) = \sum_{i=0}^{1} \sum_{a=0}^{n-1} \Pr\left(h \mid a_n = a\right) \Pr_n(i, a), \tag{4.21}$$

where $\Pr\left(h \mid a_n\right)$ is given by the Beta distribution prior equation (4.13). The expected rate is therefore:

$$\begin{aligned} \bar{h} &:= \int_0^1 h \Pr\left(h \mid \mathcal{O}_n\right) dh = \sum_{i=0}^{1} \sum_{a_n=0}^{n-1} \int_0^1 h \Pr\left(h \mid a_n\right) \Pr_n(i, a_n)\, dh \\ &= \sum_{i=0}^{1} \sum_{a_n=0}^{n-1} \frac{a_n + 1}{n + 1} \Pr_n(i, a_n) \end{aligned} \tag{4.22}$$

Explicit knowledge of the transition rate, h, is not used in the inference process described by equation (4.19). However, computing it allows us to evaluate how the observer's estimate converges to the true transition rate (see figure 4.3B).

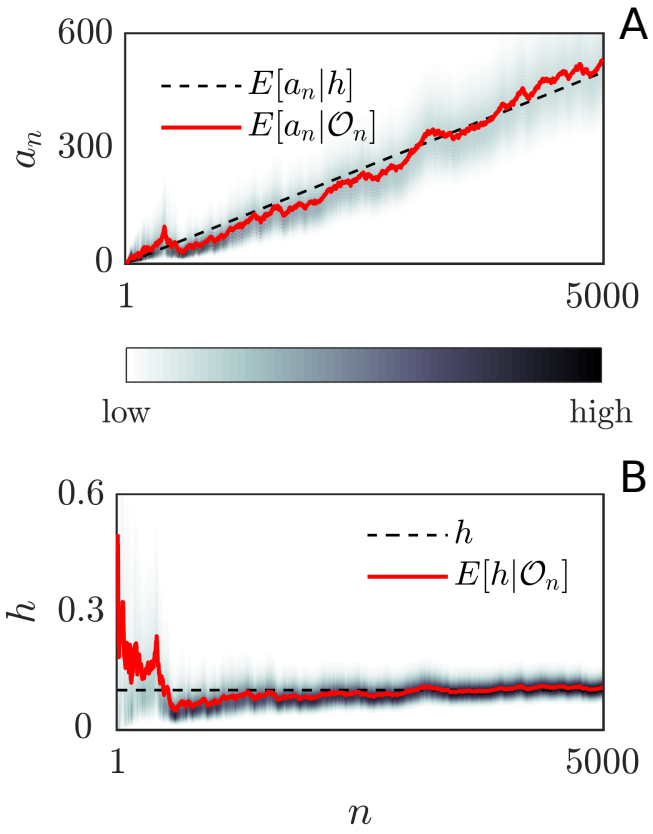

Figure 4.3: **Learning the hazard rate (image adapted from [38]). A:** Evolution of the posterior over a_n (gray scale). The posterior mean (red) converges to the expected number of change-points $h(n-1)$ (dashed line). **B:** Evolution of the posterior over the change rate h (gray scale). The posterior mean (red) converges to the true value (dashed line) and the variance diminishes with the number of observations.

Accuracy of the change-point learning model In figure 4.4, we compare the accuracy of the learning model (black curves) to that of a family of known-rate models from section 2.1.2 (green curves). Note that the learning model (4.19) reduces to the

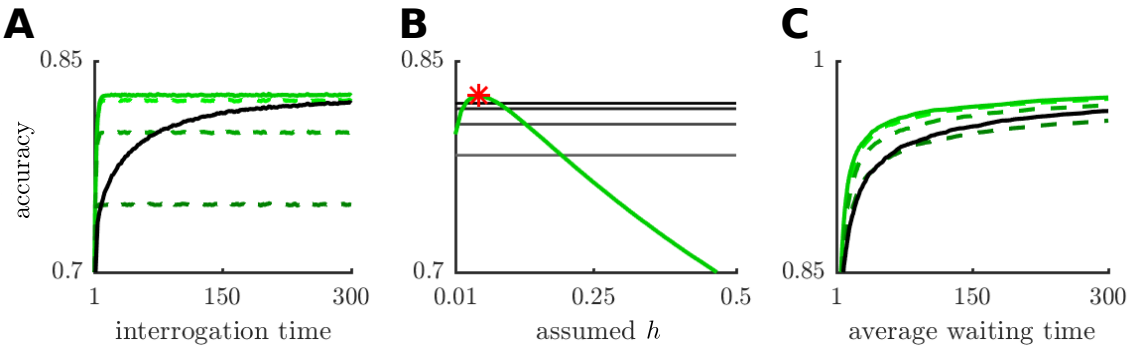

Figure 4.4: **Performance of the inference algorithm (image adapted from [38]). A:** Accuracy under the interrogation paradigm measured as the percentage of correct responses at the interrogation time. Here and in the next panel $h = 0.05$, and SNR= 1. The black curve represents the performance of an ideal observer who infers the change rate from measurements. The green curves represent the performance of observers that assume a fixed change rate (0.3, 0.15, 0.05, 0.03 from darker to lighter, see equation (2.9)). The solid green line corresponds to an observer who assumes the true rate, dashed lines to erroneous rates. **B:** The green curve represents the performance at interrogation time $n = 300$ of an observer that assumes a fixed change rate (x-axis). The red star marks the maximum of this curve, corresponding to the true change rate $h = 0.05$. The horizontal black lines represent the accuracy at times $n \in \{40, 100, 200, 300\}$ (from bottom to top) of the observer that learns the change rate. **C:** The accuracy as a function of the average threshold hitting time in the free response protocol. Here $h = 0.1$, and SNR=0.75. See appendix A for details on numerical simulations. See also figure 3 in [49].

known-rate model (2.9) when a delta-mass prior is used over the hazard rate. Thus, figure 4.4 encompasses three cases:

1. when the observer knows the true rate (delta-mass prior over the true rate h);
2. when the observer assumes a wrong rate (delta-mass prior over an erroneous h);

3. and when the observer learns the rate from measurements (flat prior over h).

Under the interrogation protocol, the observer infers the state of the environment at a fixed time. As expected, performance increases with interrogation time, and is highest if the observer uses the true rate (see figure 4.4A, also equation (2.9) above). Performance plateaus quickly when the observer assumes a fixed rate, and more slowly if the rate is learned. The performance of observers that learn the rate slowly increases toward that of observers who know the true rate. In figure 4.4B, we present the performance of the unknown-rate algorithm at 4 different times ($n \in \{40, 100, 200, 300\}$) and compare it to the asymptotic values with different assumed rates (green curves).

Note, an observer that assumes an incorrect change rate can still perform near optimally (e.g., curve for 0.03 in figure 4.4A), especially when the signal-to-noise ratio (SNR) is quite high. The SNR in the present context is the difference in means of the likelihoods divided by their common standard deviation [49]. Although this cannot be seen on the figure, the benefit of inferring the change rate is at its highest for intermediate SNR values, in which case multiple observations are needed for an accurate estimate of the present state. At very low SNR values the learning observer will not be able to substantially reduce uncertainty about the change rate, resulting in high uncertainty about the state.

In the free response protocol, the observer makes a decision when the log posterior odds ratio reaches a predefined threshold. In figure 4.4C, we present simulation results for this protocol in a format similar to figure 4.4A, with empirical performance as a function of average hitting time. Each performance level corresponds to unique log-odds threshold. Similar to the interrogation protocol (figure 4.4A), performance of the free response protocol saturates much more quickly for an observer that fixes their change rate estimate than one that infers this rate over time.

4.1.2 In an N-state environment

We next consider evidence accumulation in an environment with an arbitrary number of states, $\{1, 2, ..., M\}$, with symmetric transition probabilities, $\eta^{ij} \equiv$ constant, whenever $i \neq j$. We define $h := (N-1)\eta^{ij}$ for any $i \neq j$, so that the probability of remaining in the same state becomes $\eta^{ii} = 1 - h$, for all $i = 1, ..., M$. The symmetry

[4]This section represents a slight modification of content appearing in [38].

in transition rates means that an observer still only needs to track the total number of change-points , a_n, as in section 4.1.1.

Equations (4.6-4.7) remain valid with N possible choices, $\{1, \ldots, M\}$. When $n > 1$, the double sum in equation (4.6) simplifies to:

$$\begin{aligned}\Pr_n(i,a) =& \frac{\Pr(\mathcal{O}_{n-1})}{\Pr(\mathcal{O}_n)} f_i(o_n) \Bigg[\Pr\Big(i, a \;\Big|\; S_{n-1} = i, a_{n-1} = a\Big) \cdot \Pr_{n-1}(i,a) \\ &+ \sum_{j \in \Lambda \setminus \{i\}} \Pr\Big(i, a \;\Big|\; S_{n-1} = j, a_{n-1} = a-1\Big) \cdot \Pr_{n-1}(j, a-1) \Bigg]\end{aligned}$$

As in section 4.1.1, we have $\Pr_1(i,0) = f_i(o_1)\pi(S=i)/\Pr(o_1)$ and $\Pr_1(i,a_1) = 0$ for $a_1 \neq 0$, where $\pi(S=i)$ describes the observer's belief prior to any observations. At all future times, $n > 1$, we have at the boundaries for all $i \in \Lambda$:

$$\Pr_n(i,0) = \frac{\Pr(\mathcal{O}_{n-1})}{\Pr(\mathcal{O}_n)} f_i(o_n) \Pr\Big(S_n = i, a_n = 0 \;\Big|\; S_{n-1} = i, a_{n-1} = 0\Big) \Pr_{n-1}(i,0)$$

and,

$$\begin{aligned}\Pr_n(i,n-1) =& \frac{\Pr(\mathcal{O}_{n-1})}{\Pr(\mathcal{O}_n)} f^i(o_n) \sum_{j \neq i} \Pr\Big(i, n-1 \;\Big|\; S_{n-1} = j, a_{n-1} = n-2\Big) \\ &\times \Pr_{n-1}(j, n-2)\end{aligned}$$

Equation (4.12) remains unchanged and we still have $\Pr\Big(h \;\Big|\; S_{n-1}, a_{n-1}\Big) = \Pr\Big(h \;\Big|\; a_{n-1}\Big)$. Furthermore, assuming a Beta prior on the change rate, equation (4.13) remains valid, and equation (4.14) is replaced by:

$$\Pr\Big(S_n, a_n \;\Big|\; h, S_{n-1}, a_{n-1}\Big) = \begin{cases} 1-h & S_n = S_{n-1} \;\&\; a_n = a_{n-1} \\ h/(N-1) & S_n \neq S_{n-1} \;\&\; a_n = a_{n-1}+1 \\ 0 & \text{otherwise} \end{cases}$$

The integral from equation (4.12) gives, once again, the mean of the Beta distribution, $\hat{h}_{n-1}(a)$, defined in equations (4.16-4.17). As in section 4.1.1, $\hat{h}_{n-1}(a_{n-1})$ is a point estimate of the change rate h at time $n-1$ when the change-point count is a_{n-1}. We have,

$$\Pr\Big(S_n, a_n \;\Big|\; S_{n-1}, a_{n-1}\Big) = \begin{cases} 1-\hat{h}_{n-1}(a_n) & S_n = S_{n-1} \;\&\; a_n = a_{n-1} \\ \hat{h}_{n-1}(a_n - 1)/(N-1) & S_n \neq S_{n-1} \;\&\; a_n = a_{n-1}+1 \\ 0 & \text{otherwise} \end{cases} \tag{4.23}$$

and the main probability update equation is now:

$$\Pr_n(i,a) = \frac{\Pr(\mathcal{O}_{n-1})}{\Pr(\mathcal{O}_n)} f_i(o_n) \left[\left(1 - \hat{h}_{n-1}(a_n)\right) \cdot \Pr_{n-1}(i,a) + \frac{\hat{h}_{n-1}(a_n - 1)}{N-1} \sum_{j \in \Lambda \setminus \{i\}} \Pr_{n-1}(j, a-1) \right]$$

The observer can infer the most likely state of the environment, by computing the index that maximizes the posterior probability, marginalizing over all change-point counts,

$$\hat{\imath} = \text{argmax}_i \Pr\left(S_n = i \mid \mathcal{O}_n\right) = \text{argmax}_i \left(\sum_{a=0}^{n-1} \Pr_n(i,a) \right)$$

The observer can also compute the posterior probability $\Pr\left(h \mid \mathcal{O}_n\right)$ of the transition rate h by marginalizing over all states S_n and change-point counts a_n, as in equation (4.21). Furthermore, a point estimate of h is given by the mean of the posterior after marginalizing, as in equation (4.22).

4.2 Learning all transition rates

In this section, we depart from the framework of [4] and [51], and consider unequal transition rates between states[6]. The environment therefore behaves as in section 2.1.1. Let us briefly remind the set-up and assumptions. We consider an arbitrary number of states, N, with unknown transition rates η^{ij} between them. The switching process between the states is memoryless, so that S_n is a stationary, discrete-time Markov chain with finite state space, $\Lambda := \{1, \ldots, M\}$. We write the (unknown) transition matrix for this chain as a left stochastic matrix,

$$\boldsymbol{h} := \begin{pmatrix} \eta^{11} & \ldots & \eta^{1N} \\ \vdots & \ddots & \vdots \\ \eta^{N1} & \ldots & \eta^{NN} \end{pmatrix}$$

[5]This section represents a slight modification of content appearing in [38].

[6]This includes the possibility that some transitions are not allowed.

where $\eta^{ij} = \Pr\left(S_n = i | S_{n-1} = j\right)$, with $i, j \in \Lambda$. We denote by $\boldsymbol{h}^{\cdot i}$ the i-th column of the matrix $\boldsymbol{h}$, and similarly for other matrices. Each such column sums to 1. We define the change-point counts matrix at time n as,

$$\boldsymbol{a}_n := \begin{pmatrix} a_n^{11} & \dots & a_n^{1N} \\ \vdots & \ddots & \vdots \\ a_n^{N1} & \dots & a_n^{NN} \end{pmatrix}$$

where a_n^{ij} is the number of transitions from state j to state i up to time n. There can be a maximum of $n-1$ transitions at time n. For a fixed $n \geq 1$, all entries in $\boldsymbol{a}_n$ are nonnegative and sum to $n-1$, i.e. $\sum_i \sum_j a_n^{ij} = n-1$. As in the symmetric case, the change-point matrix at time 1 must be the zero matrix, $\boldsymbol{a}_1 = \boldsymbol{0}$.

We will show that our inference algorithm assigns positive probability only to change-point matrices that correspond to possible transition paths between the states $\{1, \dots, M\}$. Many nonnegative integer matrices with entries that sum to $n-1$ are not possible change-point matrices $\boldsymbol{a}_n$. A combinatorial argument shows that when $N = 2$, the number of possible pairs, $(S_n, \boldsymbol{a}_n)$, grows quadratically with the number of steps, n, to leading order. It can also be shown that the growth is polynomial for $N > 2$, although we do not know the growth rate in general (see figure 4.5B). An ideal observer has to assign a probability of each of these states which is much more demanding than in the symmetric rate case where the number of possible states grows linearly in n.

We next derive an iterative equation for $\Pr\left(S_n, \boldsymbol{a}_n \mid \mathcal{O}_n\right)$, the joint probability of the state S_n, and an allowable combination of the $N(N-1)$ change-point counts (off-diagonal terms of $\boldsymbol{a}_n$), and N non-change-point counts (diagonal terms of $\boldsymbol{a}_n$). The derivation is similar to the symmetric case. For $n > 1$, we first marginalize over S_{n-1} and $\boldsymbol{a}_{n-1}$,

$$\Pr\left(S_n, \boldsymbol{a}_n \mid \mathcal{O}_n\right) = \frac{1}{\Pr\left(\mathcal{O}_n\right)} \sum_{S_{n-1}, \boldsymbol{a}_{n-1}} \Pr\left(\mathcal{O}_n \mid S_n, S_{n-1}, \boldsymbol{a}_n, \boldsymbol{a}_{n-1}\right) \times \Pr\left(S_n, S_{n-1}, \boldsymbol{a}_n, \boldsymbol{a}_{n-1}\right)$$

where the sum is over all $S_{n-1} \in \{1, ..., M\}$ and possible values of the change-point matrix, $\boldsymbol{a}_{n-1}$.

Using $\Pr\left(S_n, S_{n-1}, \boldsymbol{a}_n, \boldsymbol{a}_{n-1}\right) = \Pr\left(S_n, \boldsymbol{a}_n \mid S_{n-1}, \boldsymbol{a}_{n-1}\right) \Pr\left(S_{n-1}, \boldsymbol{a}_{n-1}\right)$, and applying Bayes' rule to write

$$\Pr\left(\mathcal{O}_{n-1} \mid S_{n-1}, \boldsymbol{a}_{n-1}\right) \Pr\left(S_{n-1}, \boldsymbol{a}_{n-1}\right) = \Pr\left(S_{n-1}, \boldsymbol{a}_{n-1} \mid \mathcal{O}_{n-1}\right) \Pr\left(\mathcal{O}_{n-1}\right)$$

gives

$$\Pr\left(S_n, \boldsymbol{a}_n \mid \mathcal{O}_n\right) = \frac{\Pr\left(\mathcal{O}_{n-1}\right)}{\Pr\left(\mathcal{O}_n\right)} \Pr\left(o_n \mid S_n\right) \sum_{S_{n-1}, \boldsymbol{a}_{n-1}} \Pr\left(S_{n-1}, \boldsymbol{a}_{n-1} \mid o_{n-1}\right) \times \Pr\left(S_n, \boldsymbol{a}_n \mid S_{n-1}, \boldsymbol{a}_{n-1}\right) \tag{4.24}$$

We compute the conditional probability $\Pr\left(S_n, \boldsymbol{a}_n \mid S_{n-1}, \boldsymbol{a}_{n-1}\right)$ by marginalizing over all possible transition matrices, $\boldsymbol{h}$. To do so, we relate the probabilities of $\boldsymbol{h}$ and $\boldsymbol{a}$. Note that if the observer assumes the columns $\boldsymbol{h}^{\cdot j}$ are independent prior to any observations, then the exit rates conditioned on the change-point counts, $\boldsymbol{h}^{\cdot j} | \boldsymbol{a}_n^{\cdot j}$, are independent for all states, $j \in \Lambda$.

To motivate the derivation we first consider a single state, $j = 1$, and assume that the environmental state has been observed perfectly over $N > 1$ time steps, but the transition rates are unknown. Therefore, all $\boldsymbol{a}_n^{\cdot 1}$ are known to the observer ($1 \leq n \leq N$), but the $\boldsymbol{h}^{\cdot 1}$ are not. The state of the system at time $n+1$, given that it was in state 1 at time n, is a categorical random variable, and $\Pr\left(S_{n+1} = i \mid S_n = 1\right) = \boldsymbol{h}^{i1}$, for $1 \leq n \leq N-1$. The observed transitions $1 \mapsto i$ are independent samples from a categorical distribution with unknown parameters $\boldsymbol{h}^{\cdot 1}$.

The conjugate prior to the categorical distribution is the Dirichlet distribution, and we therefore use it as a prior on the change-point probabilities. For simplicity we again assume a flat prior over $\boldsymbol{h}^{\cdot 1}$, that is $\Pr\left(\boldsymbol{h}^{\cdot 1}\right) = \chi_L\left(\boldsymbol{h}^{\cdot 1}\right)$, where χ_L is the indicator function on the standard $(M-1)$-simplex, L.

Denote by D the sequence of states that the environment transitioned to at time $n+1$ whenever it was in state 1 at time n, for all $1 \leq n \leq N-1$. Therefore, D is a sequence of states from the set Λ. By definition,

$$\Pr\left(D \mid \boldsymbol{h}^{\cdot 1}\right) = \prod_{i=1}^{M} \left(\boldsymbol{h}^{i1}\right)^{\sum_{n=1}^{N-1} \chi(S_{n+1}=i, S_n=1)}$$

where $\chi\left(S_{n+1} = i, S_n = 1\right)$ is the indicator function, which is unity only when $S_{n+1} = i$ and $S_n = 1$ and zero otherwise. Equivalently, we can write

$$\Pr\left(\boldsymbol{a}_N^{\cdot 1} \mid \boldsymbol{h}^{\cdot 1}\right) = \prod_{i=1}^{M} \left(\boldsymbol{h}^{i1}\right)^{\boldsymbol{a}_N^{i1}}$$

since $\boldsymbol{a}_N^{i1} = \sum_{n=1}^{N-1} \chi(S_{n+1} = i, S_n = 1)$. For general $n > 1$, the posterior distribution for the transition probabilities $\boldsymbol{h}^{\cdot 1}$ given the change-point vector $\boldsymbol{a}_n^{\cdot 1}$ is then

$$\Pr\left(\boldsymbol{h}^{\cdot 1} | \boldsymbol{a}_n^{\cdot 1}\right) = \frac{\Gamma\left(\sum_{i=1}^{M}\left(a_n^{i1}+1\right)\right)}{\prod_{i=1}^{M}\Gamma\left(a_n^{i1}+1\right)} \prod_{i=1}^{M}\left(\eta^{i1}\right)^{a_n^{i1}} = dir\left(\boldsymbol{h}^{\cdot 1}; \boldsymbol{a}_n^{\cdot 1}+\mathbf{1}\right)$$

Here $\mathbf{1} = (1, ..., 1)^T$, so $\boldsymbol{a}_n^{\cdot 1} + \mathbf{1}$ should be interpreted as the vector with entries $\left(\boldsymbol{a}_n^{i1}+1\right)_{i=1}^{M}$, $\Gamma(x)$ is the gamma function, and $dir\left(\boldsymbol{h}^{\cdot 1}; \boldsymbol{a}_n^{\cdot 1}+\mathbf{1}\right)$ the probability density function of the N-dimensional Dirichlet distribution, $Dir\left(\boldsymbol{a}_n^{\cdot 1}+\mathbf{1}\right)$.

The same argument applies to all initial states, j, $j \in \{1, \ldots, M\}$. We assume that the transition rates are conditionally independent, so that

$$\Pr\left(\boldsymbol{h}|\boldsymbol{a}_n\right) = \prod_{j=1}^{M} dir(\boldsymbol{h}^{\cdot j}; \boldsymbol{a}_n^{\cdot j}+\mathbf{1}) = \prod_{j=1}^{M} \frac{\Gamma\left(\sum_{i=1}^{M}(a_n^{ij}+1)\right)}{\prod_{i=1}^{M}\Gamma((a_n^{ij}+1))} \prod_{k=1}^{M}\left(h^{kj}\right)^{a_n^{kj}} \tag{4.25}$$

Using this observation, the transition probability between two states can be computed by marginalizing over all possible transition matrices, $\boldsymbol{h}$, conditioned on $\boldsymbol{a}_{n-1}$,

$$\begin{aligned} \mathrm{P}(S_n, \boldsymbol{a}_n | S_{n-1}, \boldsymbol{a}_{n-1}) &= \int_{\mathcal{M}} \mathrm{P}(S_n, \boldsymbol{a}_n | \boldsymbol{h}, S_{n-1}, \boldsymbol{a}_{n-1}) \mathrm{P}(\boldsymbol{h} | S_{n-1}, \boldsymbol{a}_{n-1}) d\boldsymbol{h} \\ &= \int_L \cdots \int_L \mathrm{P}(S_n, \boldsymbol{a}_n | \boldsymbol{h}^{\cdot 1}, \ldots, \boldsymbol{h}^{\cdot M}, S_{n-1}, \boldsymbol{a}_{n-1}) \\ &\quad \times dir(\boldsymbol{h}^{\cdot 1}; \boldsymbol{a}_{n-1}^{\cdot 1}+1) \times \cdots \times dir(\boldsymbol{h}^{\cdot M}; \boldsymbol{a}_{n-1}^{\cdot M}+1) d\boldsymbol{h}^{\cdot 1} \cdots d\boldsymbol{h}^{\cdot M} \end{aligned} \tag{4.26}$$

where $\mathcal{M}$ represents the space of all $M \times M$ left stochastic matrices and L is the $M-1$ dimensional simplex of $\boldsymbol{h}^{\cdot j} \in [0,1]^M$ such that $\sum_{i=1}^{M} \boldsymbol{h}^{ij} = 1$.

Let $\boldsymbol{\delta}^{ij}$ be the $M \times M$ matrix containing a 1 as its ij-th entry, and 0 everywhere else. For all $i, j \in \Lambda$ we have

$$\Pr\left(S_n = i, \boldsymbol{a}_n | \boldsymbol{h}, S_{n-1} = j, \boldsymbol{a}_{n-1}\right) = \begin{cases} \eta^{ij} & \text{if } \boldsymbol{a}_n = \boldsymbol{a}_{n-1} + \boldsymbol{\delta}^{ij} \\ 0 & \text{otherwise} \end{cases} \tag{4.27}$$

Implicit in equation (4.27) is the requirement that the environment must have been in the state $S_{n-1} = j$ in order for the transition $j \mapsto i$ to have occurred between $n-1$ and n. This will ensure that the change-point matrices $\boldsymbol{a}_n$ that are assigned nonzero probability correspond to admissible paths through the states $\{1, ..., M\}$.

Applying equation (4.27), we can compute the integrals in equation (4.26) for all pairs (i,j). We let $\hat{h}^{ij}_{n-1}(\boldsymbol{a}_{n-1}) := \Pr\left(S_n = i, \boldsymbol{a}_n = \boldsymbol{a}_{n-1} + \boldsymbol{\delta}^{ij} | S_{n-1} = j, \boldsymbol{a}_{n-1}\right)$ to simplify notation, and find

$$\begin{aligned}
\hat{h}^{ij}_{n-1}(\boldsymbol{a}_{n-1}) &= \int_L \cdots \int_L \eta^{ij} \prod_{k=1}^{M} dir(\boldsymbol{h}^{\cdot k}; \boldsymbol{a}^{\cdot k}_{n-1} + 1) d\boldsymbol{h}^{\cdot 1} \cdots d\boldsymbol{h}^{\cdot M} \\
&= \int_L \eta^{ij} dir(\boldsymbol{h}^{\cdot j}; \boldsymbol{a}^{\cdot j}_{n-1} + 1) d\boldsymbol{h}^{\cdot j} \prod_{k \neq j} \int_L dir(\boldsymbol{h}^{\cdot k}; \boldsymbol{a}^{\cdot k}_{n-1} + 1) d\boldsymbol{h}^{\cdot k} \\
&= \int_L \eta^{ij} dir(\boldsymbol{h}^{\cdot j}; \boldsymbol{a}^{\cdot j}_{n-1} + 1) d\boldsymbol{h}^{\cdot j} = \frac{a^{ij}_{n-1} + 1}{M + \sum_{k=1}^{M} a^{kj}_{n-1}}
\end{aligned} \tag{4.28}$$

As in the point estimate of the rate $\hat{h}_{n-1}(a_{n-1})$ in equation (4.17), each $\hat{h}^{ij}_{n-1}(\boldsymbol{a}_{n-1})$ is a ratio containing the number of $j \mapsto i$ transitions in the numerator, and the total number of transitions out of the jth state in the denominator. Thus, the estimated transition rate $\hat{h}^{ij}_{n-1}(\boldsymbol{a}_{n-1})$ increases with the number of transitions $j \mapsto i$ in a given interval $\{1, ..., n\}$. Furthermore, each column sums to unity:

$$\sum_{i=1}^{M} \hat{h}^{ij}_{n-1}(\boldsymbol{a}_{n-1}) = \frac{\sum_{i=1}^{M} \left(a^{ij}_{n-1} + 1\right)}{N + \sum_{k=1}^{M} a^{kj}_{n-1}} = \frac{M + \sum_{i=1}^{M} a^{ij}_{n-1}}{M + \sum_{k=1}^{M} a^{kj}_{n-1}} = 1$$

so the point estimates $\hat{h}^{ij}_{n-1}(\boldsymbol{a}_{n-1})$ for the transition rates out of each state j do provide an empirical probability mass function along each column. However, as in the symmetric case, these estimates are biased toward the interior of the domain. This is a consequence of the hyper-parameters we have chosen for our prior density, $dir(\boldsymbol{h}; \boldsymbol{a}_0 + \mathbf{1})$.

Therefore, for $n > 1$, the probability update equation in the case of asymmetric transition rates (equation (4.24)) is given by,

$$\begin{aligned}
\Pr\left(S_n = i, \boldsymbol{a}_n \mid \mathcal{O}_n\right) = \frac{\Pr\left(\mathcal{O}_{n-1}\right)}{\Pr\left(\mathcal{O}_n\right)} f_i(o_n) \sum_{j=1}^{M} \hat{h}^{ij}_{n-1}(\boldsymbol{a}_n - \boldsymbol{\delta}^{ij}) \times \\
\Pr\left(S_{n-1} = j, \boldsymbol{a}_n - \boldsymbol{\delta}^{ij} \mid \mathcal{O}_{n-1}\right)
\end{aligned} \tag{4.29}$$

The point estimates of the transition rates, $\hat{h}^{ij}_{n-1}(\boldsymbol{a}_{n-1} = \boldsymbol{a}_n - \boldsymbol{\delta}^{ij})$, are defined in equation (4.28). As before, $\Pr\left(S_1 = i, \boldsymbol{a}_1 = \mathbf{0} \mid o_1\right) = f_i(o_1) P_0(S = i) / \Pr\left(o_1\right)$ and

$\Pr\left(S_1 = i, \boldsymbol{a}_1 \mid o_1\right) = 0$ for any $\boldsymbol{a}_1 \neq \mathbf{0}$. At future times, it is only possible to obtain change-point matrices $\boldsymbol{a}_n$ whose entries sum to $\sum_i \sum_j a_n^{ij} = n-1$, the change-point matrices $\boldsymbol{a}_n$ and $\boldsymbol{a}_{n-1}$ must be related as $\boldsymbol{a}_n = \boldsymbol{a}_{n-1} + \boldsymbol{\delta}^{ij}$, as noted in equation (4.27). This considerably reduces the number of terms in the sum in equation (4.29).

The observer can find the most likely state of the environment by maximizing the posterior probability after marginalizing over the change-point counts $\boldsymbol{a}_n$,

$$\hat{\imath} = \text{argmax}_i \Pr\left(S_n = i \mid \mathcal{O}_n\right) = \text{argmax}_i \left(\sum_{\boldsymbol{a}_n} \Pr\left(S_n = i, \boldsymbol{a}_n \mid \mathcal{O}_n\right) \right)$$

The transition rate matrix can also be computed by marginalizing across all possible states, S_n, and change-point count matrices, $\boldsymbol{a}_n$,

$$\Pr\left(\boldsymbol{h} \mid \mathcal{O}_n\right) = \sum_{s=1}^{M} \sum_{\boldsymbol{a}_n} \Pr\left(\boldsymbol{h} \mid \boldsymbol{a}_n\right) \Pr\left(S_n = s, \boldsymbol{a}_n \mid \mathcal{O}_n\right)$$

where $\Pr\left(\boldsymbol{h}|\boldsymbol{a}_n\right)$ is the product of probability density functions, $dir(\boldsymbol{h}^{\cdot j}; \boldsymbol{a}_n^{\cdot j} + 1)$, given in equation (4.25). The mean of this distribution is given by

$$\begin{aligned} \bar{\boldsymbol{h}} &= \int_{\mathcal{M}} \boldsymbol{h} \Pr\left(\boldsymbol{h} \mid \mathcal{O}_n\right) d\boldsymbol{h} = \sum_{s=1}^{M} \sum_{\boldsymbol{a}_n} \Pr\left(S_n = s, \boldsymbol{a}_n \mid \mathcal{O}_n\right) \int_{\mathcal{M}} \boldsymbol{h} \Pr\left(\boldsymbol{h}|\boldsymbol{a}_n\right) d\boldsymbol{h} \\ &= \sum_{s=1}^{M} \sum_{\boldsymbol{a}_n} \Pr\left(S_n = s, \boldsymbol{a}_n \mid \mathcal{O}_n\right) \boldsymbol{E}(\boldsymbol{a}_n) \end{aligned} \tag{4.30}$$

where $\boldsymbol{E}(\boldsymbol{a}_n)^{ij} = \hat{h}_n^{ij}(\boldsymbol{a}_n) = \mathbb{E}\left[\eta^{ij}|\boldsymbol{a}_n\right]$ defined in equation (4.28), is a conditional expectation over each possible change-point matrix $\boldsymbol{a}_n$.

Equation (4.29) is easier to interpret when $N = 2$. Using equation (4.28), we find

$$\hat{h}_{n-1}^{21}(\boldsymbol{a}_{n-1}) = \frac{a_{n-1}^{21} + 1}{2 + a_{n-1}^{21} + a_{n-1}^{11}}, \quad \hat{h}_{n-1}^{12}(\boldsymbol{a}_{n-1}) = \frac{a_{n-1}^{12} + 1}{2 + a_{n-1}^{12} + a_{n-1}^{22}}$$

and we can express $\hat{h}_{n-1}^{11}(\boldsymbol{a}_{n-1}) = 1 - \hat{h}_{n-1}^{21}(\boldsymbol{a}_{n-1})$ and $\hat{h}_{n-1}^{22}(\boldsymbol{a}_{n-1}) = 1 - \hat{h}_{n-1}^{12}(\boldsymbol{a}_{n-1})$.

Expanding the sum in equation (4.29), we have

$$\Pr\left(S_n = 1, \boldsymbol{a}_n \mid \mathcal{O}_n\right) = \frac{\Pr\left(\mathcal{O}_{n-1}\right)}{\Pr\left(\mathcal{O}_n\right)} f_1(o_n) \left[\hat{h}^{11}_{n-1}(\boldsymbol{a}_n - \boldsymbol{\delta}^{11}) \Pr\left(S_{n-1} = 1, \boldsymbol{a}_n - \boldsymbol{\delta}^{11} \mid \mathcal{O}_{n-1}\right) + \hat{h}^{12}_{n-1}(\boldsymbol{a}_n - \boldsymbol{\delta}^{12}) \Pr\left(S_{n-1} = 2, \boldsymbol{a}_n - \boldsymbol{\delta}^{12} \mid \mathcal{O}_{n-1}\right)\right] \tag{4.31a}$$

$$\Pr\left(S_n = 2, \boldsymbol{a}_n \mid \mathcal{O}_n\right) = \frac{\Pr\left(\mathcal{O}_{n-1}\right)}{\Pr\left(\mathcal{O}_n\right)} f_2(o_n) \left[\hat{h}^{22}_{n-1}(\boldsymbol{a}_n - \boldsymbol{\delta}^{22}) \Pr\left(S_{n-1} = 2, \boldsymbol{a}_n - \boldsymbol{\delta}^{22} \mid \mathcal{O}_{n-1}\right) + \hat{h}^{21}_{n-1}(\boldsymbol{a}_n - \boldsymbol{\delta}^{21}) \Pr\left(S_{n-1} = 1, \boldsymbol{a}_n - \boldsymbol{\delta}^{21} \mid \mathcal{O}_{n-1}\right)\right] \tag{4.31b}$$

The boundary and initial conditions will be given as above, and the mean inferred transition matrix is given by equation (4.30). Importantly, the inference process described by Eqs. (4.31) allows for both asymmetric change-point matrices, $\boldsymbol{a}_n$, and inferred transition rate matrices $\boldsymbol{E}(\boldsymbol{a}_n)$, unlike the process in equation (4.19). However, the variance of the posteriors over the rates will decrease more slowly, as fewer transitions out of each particular state will be observed.

This algorithm can be used to infer unequal transition rates as shown in figure 4.5: Panels C through E show that the mode of the joint posterior distribution, $\Pr\left(\eta^{21}, \eta^{12} \mid \mathcal{O}_n\right)$, approaches the correct rates, while its variance decreases. As in section 4.1.1 we conjecture that this joint density does not converge to a point mass at the true rate values unless the SNR is infinite.

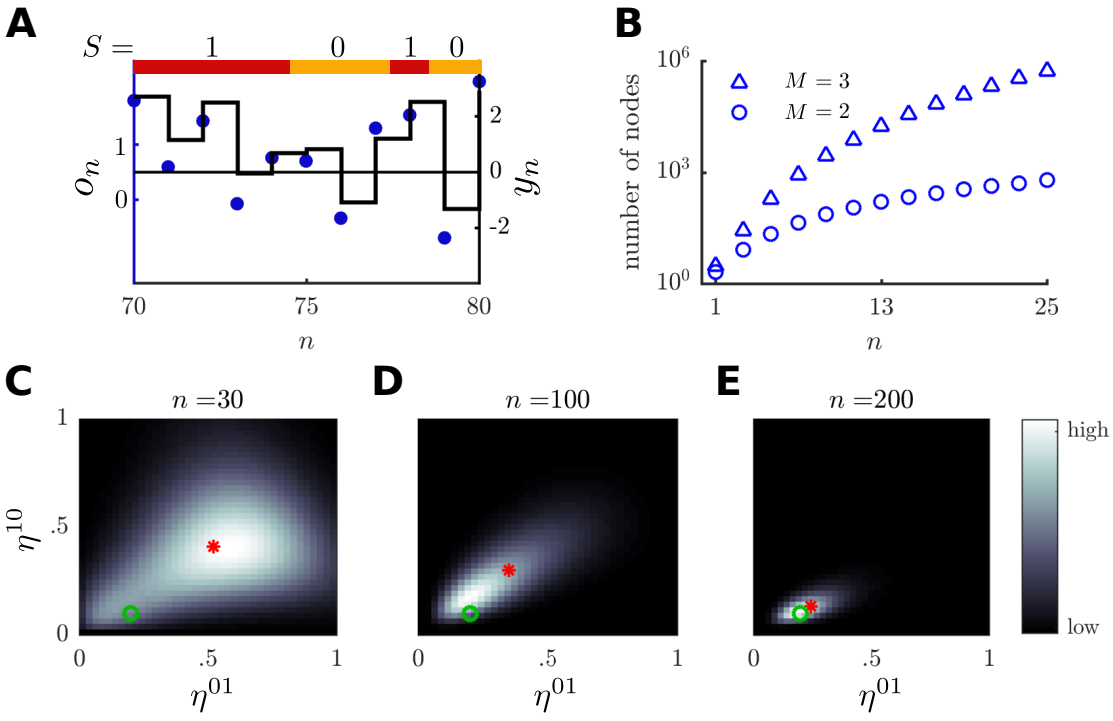

Figure 4.5: **Evidence accumulation and change rates inference in a 2-state asymmetric system (image adapted from [38]). A:** Sample path (color bar, top) of the environment between time steps 70 and 80 (same simulation as in panels C-E) with corresponding observations (blue dots), and log posterior odds ratio (black step function). Here and in panels C-E, $\left(\eta^{01}, \eta^{10}\right) = (0.2, 0.1)$, SNR= 1.4. **B:** The number of allowable change-point count matrices as a function of observation number, n, for $M = 2$ (blue circles), and $M = 3$ (blue triangles). **C-E:** Color plots (gray scale) of the joint density, $\Pr\left(\eta^{01}, \eta^{10} \mid \mathcal{O}_n\right)$, with mean value (red star) approaching the true transition rates (green circle), as number of observations increases.

4.3 Summary

We discovered two ways of expressing an ideal observer model of decision making in symmetrically changing environments. One possibility is to evolve the joint posterior probability over the state and the hazard rate. The other is to consider the joint posterior over the state and change-point count. Focusing on the latter strategy, we observed that the hazard rate could be learned in the appropriate parameter regimes, and that this learning was accompanied by an increase in choice accuracy. Thus, as the number of observations increases, the learning model outperforms more and more ideal observer models who use a wrong hazard rate value. Finally, we extended the change-point count strategy to asymmetrically changing environments and found the optimal inference algorithm heavy to implement. More specifically, the support of the joint posterior over state and change-point counts in the case of asymmetric change rates becomes too big too quickly. This is due to the combinatorial explosion in the number of possible change-point counts matrices as time progresses.

Chapter 5

Learning the hazard rate in continuous time

Some tasks are more naturally modeled in continuous time. This is the case, for instance, of the dots reversal tasks and the dynamic clicks task, described in chapter 1. In the present chapter, we derive the continuum limits of the ideal observer models from chapter 4. Thus, the environment is now a continuous time 2-state Markov chain with symmetric exit rate from each state, h.

5.1 Joint inference on state and hazard rate

Our discrete time equations (4.4) describing joint inference on state and hazard rate are equivalent to a family of filtering equations like (2.2a) indexed by $h \in H$, with the additional normalization constraint (4.5). This implies that all the developments leading to the extended SPRT equation (2.9) are valid. Thus, with the normalization (4.5) enforced, we get,

$$y_n^h = I_n + y_{n-1}^h + \log \frac{h \cdot \exp\left(-y_{n-1}^h\right) + (1-h)}{h \cdot \exp\left(y_{n-1}^h\right) + (1-h)} \qquad \forall (h,n) \in H \times \mathbb{N}_{>1} \tag{5.1a}$$

$$y_1^h = I_1 + y_0^h \qquad \forall h \in H \tag{5.1b}$$

where the superscript h is just an indexing symbol. Note how the observational input I_n and the initial condition $y_0^h \coloneqq \log\big(\pi(S=1)/\pi(S=0)\big)$ are the same for all h. The analogy with section 2.1.2 follows through the continuum limit (2.18) as well.

Hence, the evidence accumulated by an ideal observer who jointly infers the state and hazard rate may be characterized by the following family of SDEs indexed by $h \in H$,

$$dy_t^h = m_t dt + D_t dW_t - 2h \cdot \sinh\left(y_t^h\right) dt \tag{5.2}$$

with deterministic initial conditions $y_0^h = 0,\ \forall h \in H$. Importantly, in the above equation, the realization of the noise, $D_t dW_t$, is the *same* across the whole family.

5.2 Joint inference on state and change-point count

In the continuous time setting, the change-point counting process $\{a_t\}_{t\geq 0}$ associated with the environment is now a Poisson process with birth rate h. We will derive here the continuum limit of the inference process (4.19), using the same sampled-time approximation method presented in section 2.2.2.

Some care must be taken regarding the correspondence between the discrete time and continuous time versions of the hazard rate. For the sake of clarity, let us denote by $h \in [0, \infty)$ the hazard rate in continuous time, and by $h^{\Delta t} \in [0, 1]$ its discrete time counterpart. In section 4.1.1, we used a Beta prior on $h^{\Delta t}$ with shape hyper-parameters $a_0, b_0 > 0$,

$$\pi\left(h^{\Delta t}\right) = \tilde{\beta}\left(h^{\Delta t}; a_0, b_0\right) \coloneqq \frac{\left(h^{\Delta t}\right)^{a_0 - 1}\left(1 - h^{\Delta t}\right)^{b_0 - 1}}{B(a_0, b_0)} \tag{5.3}$$

where the Beta function B is defined in equation (5.5a) below. Here, in order to select the conjugate prior to the Poisson likelihood (because of the change-point counting process), we assume a Gamma prior on h with respective shape and rate hyper-parameters $\alpha, \beta > 0$,

$$\pi(h) = \tilde{\gamma}(h; \alpha, \beta) \coloneqq \frac{\beta^\alpha h^{\alpha - 1} \exp(-\beta h)}{\Gamma(\alpha)} \tag{5.4}$$

with the Gamma function Γ defined in equation (5.5b) below. In the sampled-time approximation, we set,

$$h^{\Delta t} \coloneqq h \Delta t$$

The question is, how should (a_0, b_0) relate to (α, β) in order for the limit $\Delta t \to 0$ to make sense? Thanks to the following lemma, the answer is:

$$\begin{aligned} a_0 &\coloneqq \alpha \\ b_0 &\coloneqq \beta / \Delta t \end{aligned}$$

Lemma 5.2.1. *Let $h \sim Gamma(\alpha, \beta)$ for some $\alpha, \beta > 0$, and define the following family of Beta-distributed random variables: $h_n \sim Beta(\alpha, n\beta)$. Then, the sequence $\{nh_n\}_{n\in\mathbb{N}}$ converges to h in distribution, as $n \to \infty$.*

Proof. We build up on the online proof from [36]. We start by establishing some notation. The symbols B and Γ will represent, respectively, the Beta and Gamma functions, restricted to the positive reals:

$$B(x, y) \coloneqq \int_0^1 t^{x-1}(1-t)^{y-1}dt \qquad \forall x, y > 0 \tag{5.5a}$$

$$\Gamma(z) \coloneqq \int_0^\infty x^{z-1}\exp(-x)dx \qquad \forall z > 0 \tag{5.5b}$$

For fixed positive reals $\alpha, \beta > 0$, define the two sequences of independent random variables $\{X_n\}_{n\in\mathbb{N}}, \{Y_n\}_{n\in\mathbb{N}}$, by $X_n \sim Beta(\alpha, \beta n)$, and $Y_n \coloneqq \beta n X_n$. For any $s \in \mathbb{N}$, standard probability theory tells us that, for all $n \in \mathbb{N}$:

$$\mathbb{E}\left[X_n^s\right] = \prod_{r=0}^{s-1} \frac{\alpha + r}{\alpha + \beta n + r} = \frac{B(\alpha + s, \beta n)}{B(\alpha, \beta n)} = \frac{\Gamma(\alpha + s)\Gamma(\alpha + \beta n)}{\Gamma(\alpha + s + \beta n)\Gamma(\alpha)} \tag{5.6}$$

Using the identity $z\Gamma(z) = \Gamma(z + 1)$ for any positive real $z > 0$, induction yields, $\Gamma(z + m) = (z + m - 1)\cdots(z + 1)z\Gamma(z)$, for all $m \in \mathbb{N}$. Thus, with $z \coloneqq \alpha + \beta n$, we get:

$$\begin{aligned}\frac{\Gamma(\alpha + \beta n)}{\Gamma(\alpha + s + \beta n)} &= \frac{1}{(\alpha + \beta n)(\alpha + \beta n + 1)\cdots(\alpha + \beta n + s - 1)} \\ &= (\beta n)^{-s}\left[\left(1 + \frac{\alpha}{\beta n}\right)\left(1 + \frac{\alpha + 1}{\beta n}\right)\cdots\left(1 + \frac{\alpha + s - 1}{\beta n}\right)\right]^{-1}\end{aligned} \tag{5.7}$$

Since $\mathbb{E}\left[Y_n^s\right] = (\beta n)^s\mathbb{E}\left[X_n^s\right]$ by construction, we can use equations (5.6) and (5.7) to conclude that:

$$\lim_{n\to\infty} \mathbb{E}\left[Y_n^s\right] = \lim_{n\to\infty} (\beta n)^s\mathbb{E}\left[X_n^s\right] = \Gamma(\alpha + s)/\Gamma(\alpha) \tag{5.8}$$

On the other hand, if Z is a Gamma-distributed random variabe with hyper-parameters $\alpha, 1 > 0$, then $\mathbb{E}\left[Z^s\right] = (\alpha + s - 1)\cdots\alpha = \Gamma(\alpha + s)/\Gamma(\alpha)$, for all $s \in \mathbb{N}$. So, we have proven that all the moments of the sequence $\{Y_n\}_{n\in\mathbb{N}}$ converge to the corresponding moments of Z. This implies that Y_n converges to $Z \sim Gamma(\alpha, 1)$ in distribution. By the scaling property of the Gamma distribution, this implies, in turn, that the limiting distribution of Y_n/β is $Gamma(\alpha, \beta)$, which proves our lemma. □

Thus, the expectation $\hat{h}_n(a)$ from equation (4.16) with general hyper-parameters α, β must be adapted to:

$$\hat{h}^{\Delta t}_{t_n}(a) \coloneqq \frac{a+\alpha}{n-1+\alpha+\beta/\Delta t} = \Delta t \frac{a+\alpha}{t_{n-1}+\alpha\Delta t+\beta} \tag{5.9}$$

The domains of each variable in equation (5.9) are as follows:

$$\alpha, \beta > 0 \qquad 0 \leq a \leq n-1 \qquad t_{n-1} \coloneqq (n-1)\Delta t \tag{5.10a}$$

$$(\Delta t, n) \in \left\{ (x,y) \in (0,\infty) \times \mathbb{N} \;\middle|\; xy \in [0,T] \right\} \tag{5.10b}$$

Equipped with the correspondence (5.9), we may now embed equation (4.19) in continuous time:

$$\begin{aligned} \Pr_{t_n}(i,a) = &\frac{\Pr\left(\mathcal{O}_{t_{n-1}}\right)}{\Pr\left(\mathcal{O}_{t_n}\right)} f_i^{\Delta t}(o_{t_n}) \left[\left(1 - \hat{h}^{\Delta t}_{t_{n-1}}(a)\right) \Pr_{t_{n-1}}(i,a) \right. \\ &\left. + \hat{h}^{\Delta t}_{t_{n-1}}(a-1) \Pr_{t_{n-1}}(1-i, a-1) \right] \end{aligned} \tag{5.11a}$$

Equation (5.11) is only valid for $i \in \Lambda, n > 1$, and $0 < a < n-1$. Just as in equations (4.9)-(4.11), the following initial and boundary conditions hold for all $i \in \Lambda$. At $n = 1$,

$$\begin{gathered} \Pr_{t_1}(i,0) = \frac{1}{\Pr\left(\mathcal{O}_{t_1}\right)} f_i^{\Delta t}(o_{t_1}) \pi(S=i) \\ a > 0 \implies \Pr_{t_1}(i,a) = 0 \end{gathered} \tag{5.11b}$$

and for $n > 1$, the boundary values for $a \in \{0, n-1\}$ yield:

$$\Pr_{t_n}(i,0) = \frac{\Pr\left(\mathcal{O}_{t_{n-1}}\right)}{\Pr\left(\mathcal{O}_{t_n}\right)} f_i^{\Delta t}(o_{t_n}) \left(1 - \hat{h}^{\Delta t}_{t_{n-1}}(0)\right) \Pr_{t_{n-1}}(i,0) \tag{5.11c}$$

$$\Pr_{t_n}(i,n-1) = \frac{\Pr\left(\mathcal{O}_{t_{n-1}}\right)}{\Pr\left(\mathcal{O}_{t_n}\right)} f_i^{\Delta t}(o_{t_n}) \hat{h}^{\Delta t}_{t_{n-1}}(n-2) \Pr_{t_{n-1}}(1-i, n-2) \tag{5.11d}$$

The process of deriving a SDE for the log posterior odds ratio, as we did in the previous continuum limit procedures, is now hampered by the fact that equations (5.11) are coupled along the a-dimension. Instead, we derive a system of SDEs for the log-posteriors themselves. As a first step, we divide both sides of equation (5.11a) by $\Pr_{t_{n-1}}(i,a)$, yielding,

$$\begin{aligned} \frac{\Pr_{t_n}(i,a)}{\Pr_{t_{n-1}}(i,a)} = &\frac{\Pr\left(\mathcal{O}_{t_{n-1}}\right)}{\Pr\left(\mathcal{O}_{t_n}\right)} f_i^{\Delta t}(o_{t_n}) \left[\left(1 - \hat{h}^{\Delta t}_{t_{n-1}}(a)\right) \right. \\ &\left. + \hat{h}^{\Delta t}_{t_{n-1}}(a-1) \cdot \frac{\Pr_{t_{n-1}}(1-i, a-1)}{\Pr_{t_{n-1}}(i,a)} \right] \end{aligned} \tag{5.12a}$$

for $i \in \Lambda, n > 1$, and $0 < a < n-1$. If we set $\Pr_{t_0}(i,0) \coloneqq \pi(S=i)$ for all $i \in \Lambda$, equation (5.11b) becomes:

$$\begin{gathered}\frac{\Pr_{t_1}(i,0)}{\pi(S=i)} = \frac{1}{\Pr(\mathcal{O}_{t_1})} f_i^{\Delta t}(o_{t_1}) \\ a > 0 \implies \Pr_{t_1}(i,a) = 0\end{gathered} \tag{5.12b}$$

Equation (5.11c) becomes, for $n > 1$:

$$\frac{\Pr_{t_n}(i,0)}{\Pr_{t_{n-1}}(i,0)} = \frac{\Pr(\mathcal{O}_{t_{n-1}})}{\Pr(\mathcal{O}_{t_n})} f_i^{\Delta t}(o_{t_n}) \left(1 - \hat{h}_{t_{n-1}}^{\Delta t}(0)\right) \tag{5.12c}$$

As for equation (5.11d), we must have:

$$\frac{\Pr_{t_n}(i,n-1)}{\Pr_{t_{n-1}}(i,n-1)} = \infty \tag{5.12d}$$

For the ranges made explicit in (5.10), we define,

$$\begin{aligned}\mathrm{x}_{i,a}(t_n) &\coloneqq \log \Pr_{t_n}(i,a) \\ \Delta\mathrm{x}_{i,a}(t_n) &\coloneqq \mathrm{x}_{i,a}(t_n) - \mathrm{x}_{i,a}(t_{n-1})\end{aligned}$$

so that equations (5.12) become,

$$\begin{aligned}\Delta\mathrm{x}_{i,a}(t_n) = C + \log f_i^{\Delta t}(o_{t_n}) + \log \Bigg[&\left(1 - \hat{h}_{t_{n-1}}^{\Delta t}(a)\right) \\ &+ \hat{h}_{t_{n-1}}^{\Delta t}(a-1) \cdot \frac{\exp\left(\mathrm{x}_{1-i,a-1}(t_{n-1})\right)}{\exp\left(\mathrm{x}_{i,a}(t_{n-1})\right)} \Bigg]\end{aligned} \tag{5.13a}$$

$$\Delta\mathrm{x}_{i,0}(t_1) = C + \log f_i^{\Delta t}(o_{t_1}) \qquad (\mathrm{x}_{i,a}(t_1) \text{ undefined for } a > 0) \tag{5.13b}$$

$$\Delta\mathrm{x}_{i,0}(t_n) = C + \log f_i^{\Delta t}(o_{t_n}) + \log\left(1 - \hat{h}_{t_{n-1}}^{\Delta t}(0)\right) \tag{5.13c}$$

$$\Delta\mathrm{x}_{i,n-1}(t_n) = \infty \tag{5.13d}$$

with $C = \log\left(1/\Pr(\mathcal{O}_{t_n})\right)$ in (5.13b) and $C = \log\left(\Pr(\mathcal{O}_{t_{n-1}})/\Pr(\mathcal{O}_{t_n})\right)$ in (5.13a) and (5.13c). Our next step is to add two assumptions:

1. The number of change-points is bounded above, irrespective of the time step Δt. So, there is an upper bound $A \in \mathbb{N}$ such that for all $\Delta t > 0$ and all $n \in \mathbb{N}$, $a < A$.

2. The ratio of exponentials in (5.13a) is bounded above, irrespective of the time step Δt.

The first assumption amounts to truncating the tail of the poisson distribution with mean hT, since this is the distribution of the change-point count variable a_T. Note that this, in turn, will prevent the system of SDEs (5.15a) from being infinite-dimensional. Finally, this first assumption is necessary to ensure that,

$$\hat{h}^{\Delta t}_{t_n}(a) = O(\Delta t)$$

The second assumption amounts to assuming that the ratio of posterior probabilities $\Pr_{t_{n-1}}(1-i, a-1)/\Pr_{t_{n-1}}(i, a)$ doesn't become too big. This is a reasonable assumption in low-to-intermediate SNR regimes[1], as the observer never totally discards any of the (S_t, a_t) combinations[2]. Both assumptions above then allow us to Taylor-expand the logarithm term in equations (5.13a) and (5.13c) about 1. Equation (5.13a) becomes,

$$\Delta \mathrm{x}_{i,a}(t_n) = C + \log f_i^{\Delta t}(o_{t_n}) + \Delta t \left[-\frac{\hat{h}^{\Delta t}_{t_{n-1}}(a)}{\Delta t} + \frac{\hat{h}^{\Delta t}_{t_{n-1}}(a-1)}{\Delta t} \cdot \frac{\exp\left(\mathrm{x}_{1-i,a-1}(t_{n-1})\right)}{\exp\left(\mathrm{x}_{i,a}(t_{n-1})\right)} \right] + o(\Delta t) \tag{5.14a}$$

and equation (5.13c) becomes:

$$\Delta \mathrm{x}_{i,0}(t_n) = C + \log f_i^{\Delta t}(o_{t_n}) - \Delta t \frac{\hat{h}^{\Delta t}_{t_{n-1}}(0)}{\Delta t} + o(\Delta t) \tag{5.14b}$$

The final step of our continuum limit derivation comprises three parts. First, at any given time $t_n \in [0, T]$, the constant C is independent of both the state and change-point count. Thus, it is not required to infer the most probable state at interrogation time, and we dismiss it from our upcoming SDEs. Second, the $\log f_i^{\Delta t}(o_{t_n})$-term from all equations above may be treated as I_k in section 2.2.1. Namely, it produces the drift and diffusion terms in equation (5.15a) below. Third, the remaining terms of equation (5.14) yield an ODE upon dividing through by Δt and letting $\Delta t \to 0$. As

[1] Note that this assumption also requires a prior π close to uniform.

[2] Except for large $a_t > A$, as per assumption 1, of course.

final result, we obtain the following system of SDEs,

$$d\mathrm{x}_{i,a}(t) = m_{i,t}dt + D_{i,t}dW_{i,t} + \left(\frac{a+\alpha-1}{t+\beta}\cdot\frac{\exp\left(\mathrm{x}_{1-i,a-1}(t)\right)}{\exp\left(\mathrm{x}_{i,a}(t)\right)} - \frac{a+\alpha}{t+\beta}\right)dt \tag{5.15a}$$

where $m_{i,t} = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\mathrm{E}_o[\ln f_i^{\Delta t}(o)|S_t]$ and $D_{i,t}$ satisfies $\langle D_{i,t}W_{i,t}D_{j,t}W_{j,t}\rangle = \Sigma_t^{ij}\cdot t$, with $\Sigma_t^{ij} = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\mathrm{Cov}_o[\ln f_i^{\Delta t}(o), \ln f_j^{\Delta t}(o)|S_t]$ for $i,j\in\Lambda$. The system (5.15a) is valid for $i\in\Lambda, t\in[0,T]$, and $0<a<A$. When $t>0$ and $a=0$, equation (5.14b) gives us instead:

$$d\mathrm{x}_{i,0}(t) = m_{i,t}dt + D_{i,t}dW_{i,t} - \frac{a+\alpha}{t+\beta}dt \tag{5.15b}$$

Equation (5.13d) doesn't cause any problem because of the truncation $a_t < A$. Finally, at $t=0$, the prior belief over the states prevails, with all the probability mass allocated to the 0-change-point nodes: $\mathrm{x}_{i,0}(0) = \log\pi(S=i)$.

Lastly, it is possible to implement the change of variables, $\mathrm{Pr}_t(i,a) \coloneqq \exp\left(\mathrm{x}_{i,a}(t)\right)$, in order to derive evolution equations for the approximate[3] posteriors, $\mathrm{Pr}_t(i,a)$. Itô's formula [21] applied to equations (5.15) thus gives:

$$d\mathrm{Pr}_t(i,a) = \mathrm{Pr}_t(i,a)\left[\left(m_{i,t}+\frac{1}{2}\right)dt + D_{i,t}dW_{i,t}\right] + \left[\frac{a+\alpha-1}{t+\beta}\mathrm{Pr}_t(1-i,a-1) - \frac{a+\alpha}{t+\beta}\mathrm{Pr}_t(i,a)\right]dt \tag{5.16a}$$

$$d\mathrm{Pr}_t(i,a) = \mathrm{Pr}_t(i,a)\left[\left(m_{i,t}+\frac{1}{2}-\frac{a+\alpha}{t+\beta}\right)dt + D_{i,t}dW_{i,t}\right] \tag{5.16b}$$

5.3 Learning in the dynamic clicks task

We derive below, the equivalent of the system of SDEs (5.15) for an ideal observer model in the dynamic clicks task. Just as in section 2.2.5, since the evidence is pulsatile, the continuum limit will give rise to a system of jump ODEs. However, as we will see, great care must be taken in this limiting procedure.

[3]Recall from section 2.2.1 how these continuum limits only represent ensemble statistics of the log posterior odds ratio.

Let us first remind our reader that with pulsatile evidence, t_n denotes the time bin $\big[n\Delta t,\ (n+1)\Delta t\big)$, for some $(n, \Delta t) \in \mathbb{N} \times (0, \infty)$ satisfying $(n+1)\Delta t \leq T$. Observation $o_{t_n} \in \Xi = \{\mathtt{10}, \mathtt{01}, \mathtt{00}\}$, is then the click count for each stream over time bin t_n, while the state $S_{t_n} \in \Lambda = \{0, 1\}$ is defined as the n-th step of a discrete time Markov chain with symmetric cross-state transition probability $h\Delta t$. Every time we mention the limit $\Delta t \to 0$, it is understood that we simultaneously let $n \to \infty$ in such a way that $n\Delta t \to t$ for some $t \in [0, T]$. Using such sampled-time approximation of the continuous processes $\{S_t\}_{t\geq 0}, \{o_t\}_{t\geq 0}$, all the derivations from section 5.2 leading to equations (5.13) and (5.14) remain valid, with the caveat that the likelihood functions should now be interpreted according to equations (2.19).

The two recurring terms in equations (5.13),

$$C + \log f_i^{\Delta t}(o_{t_n}) \tag{5.17}$$

are problematic when trying to divide through by Δt and let $\Delta t \to 0$. More specifically, the difficulty appears when $o_{t_n} \in \{\mathtt{10}, \mathtt{01}\}$ is a single-click observation, at which point the likelihoods $f_i^{\Delta t}(o_{t_n})$ scale linearly with Δt $(i = 0, 1)$. Dividing through by Δt generates a singular term of the form $\log(\Delta t)/\Delta t$. We distinguish two cases for a fixed time bin t_n:

Case 1: $o_{t_n} = \mathtt{00}$ is a no-click observation

Case 2: $o_{t_n} \in \{\mathtt{10}, \mathtt{01}\}$ is a single-click observation

In section 5.3.1 we explain why $(C + \log f_i^{\Delta t}(o_{t_n}))/\Delta t \to 0$ as $\Delta t \to 0$ for case 1. In section 5.3.2, we show how the singularity problem from case 2 can be circumvented.

5.3.1 No-click time window (Case 1)

Our program is to show that that when $o_{t_n} = \mathtt{00}$, the following equation holds for any $\Delta t > 0, n > 1$:

$$\log \frac{\Pr\left(\mathcal{O}_{t_{n-1}}\right)}{\Pr\left(\mathcal{O}_{t_n}\right)} + \log f_i^{\Delta t}(o_{t_n}) = 0 \tag{5.18}$$

It is clear how this implies:

$$\lim_{\Delta t \to 0} \frac{C + \log f_i^{\Delta t}(o_{t_n})}{\Delta t} = 0$$

Our strategy is to show that,

$$\frac{\Pr\left(\mathcal{O}_{t_{n-1}}\right)}{\Pr\left(\mathcal{O}_{t_n}\right)} = \frac{1}{f_i^{\Delta t}(o_{t_n})} \tag{5.19}$$

To ease notation, we will equate n with the time index $t_n \in (0, T)$, and denote by $S_{1:n} \coloneqq (S_1, \ldots, S_n)$, a realization of the first n steps of the environment. It will also prove useful to define the set $\mathscr{S}_{n-1}^i$ $(i \in \Lambda, n > 1)$ as the set of all paths $S_{1:n-1}$ ending in state $S_{n-1} = i$. Thus, $\left\{\mathscr{S}_{n-1}^i\right\}_{i\in\{0,1\}}$ is a 2-element partition of the set of all environmental paths of length $n-1$.

Using the law of total probability and the conditional independence of the observations on the states, we have,

$$\frac{\Pr\left(\mathcal{O}_{n-1}\right)}{\Pr\left(\mathcal{O}_n\right)} = \frac{\sum_{S_{1:n-1}} \Pr\left(S_{1:n-1}\right) \prod_{j=1}^{n-1} f_{S_j}^{\Delta t}(o_j)}{\sum_{S_{1:n}} \Pr\left(S_{1:n}\right) \prod_{j=1}^{n} f_{S_j}^{\Delta t}(o_j)} \tag{5.20}$$

where the sums are over all corresponding finite paths of the environment chain. Observe that the Markov property gives us,

$$P(S_{1:n}) = P(S_n|S_{1:n-1})P(S_{1:n-1}) = P(S_n|S_{n-1})P(S_{1:n-1}), \quad \forall n > 1$$

Hence, we may rewrite (5.20) as:

$$\frac{P(\mathcal{O}_{n-1})}{P(\mathcal{O}_n)} = \frac{\sum_{S_{1:n-1}} P(S_{1:n-1}) \prod_{j=1}^{n-1} f_{S_j}^{\Delta t}(o_j)}{\sum_{S_{1:n}} f_{S_n}^{\Delta t}(o_n) P(S_n|S_{n-1}) P(S_{1:n-1}) \prod_{j=1}^{n-1} f_{S_j}^{\Delta t}(o_j)} \tag{5.21}$$

We now introduce yet another piece of notation to keep the equations as light as possible:

$$\mathcal{F}_{n-1}^i \coloneqq \sum_{S_{1:n-1}\in\mathscr{S}_{n-1}^i} P(S_{1:n-1}) \prod_{j=1}^{n-1} f_{S_j}^{\Delta t}(o_j). \tag{5.22}$$

Our next step is to rewrite the right-hand side of equation (5.21) as:

$$\frac{\mathcal{F}_{n-1}^1 + \mathcal{F}_{n-1}^0}{\sum_{S_n\in\Lambda} \left[f_{S_n}^{\Delta t}(o_n) P(S_n|S_{n-1}=1) \mathcal{F}_{n-1}^1 + f_{S_n}^{\Delta t}(o_n) P(S_n|S_{n-1}=0) \mathcal{F}_{n-1}^0 \right]} \tag{5.23}$$

Since the probability of a transition $(S_n \neq S_{n-1})$ within a single time step is[4] $h \cdot \Delta t$, expression (5.23) simplifies to,

$$\frac{P(\mathcal{O}_{n-1})}{P(\mathcal{O}_n)} = \frac{\mathcal{F}_{n-1}^1 + \mathcal{F}_{n-1}^0}{f_1^{\Delta t}(o_n)\mathcal{F}_{n-1}^1 + f_o^{\Delta t}(o_n)\mathcal{F}_{n-1}^0 + K}, \tag{5.24}$$

[4]We are aware that h is unknown in our current learning model. However, only the "notion" of a transition rate will be used below, never the value of h by itself.

with K defined by,

$$K \coloneqq h \cdot \Delta t \cdot \left[f_1^{\Delta t}(o_n)\mathcal{F}_n^0 + f_0^{\Delta t}(o_n)\mathcal{F}_n^1 - f_1^{\Delta t}(o_n)\mathcal{F}_n^1 - f_0^{\Delta t}(o_n)\mathcal{F}_n^0 \right] \tag{5.25}$$

When $o_n = \mathtt{oo}$, equation (2.19) tells us that:

$$f_1^{\Delta t}(\mathtt{oo}) = f_0^{\Delta t}(\mathtt{oo})$$

This has several important consequences. The first one is that K vanishes as the likelihood terms appearing in its definition are equal. The second consequence is that the likelihoods in the denominator of equation (5.24) may be factored out, thereby proving equation (5.19). Equation (5.18) follows naturally from this.

5.3.2 One-click time window (Case 2)

In this second case, we separate the likelihood term into a singular and non-singular part:

$$\log f_i^{\Delta t}(o_{t_n}) = \log(\Delta t) + \log(\lambda_{\text{case}})$$

This suggests the following piecewise definitions of the new functions w and F^i, for $i \in \Lambda, n > 1$, and $0 < a < \min\{n-1, A\}$ ($A \in \mathbb{N}$ large):

$$w(t_n) \coloneqq \begin{cases} 0 & \text{if } o_{t_n} = \mathtt{oo} \\ \log \dfrac{P(\mathcal{O}_{t_{n-1}})}{P(\mathcal{O}_{t_n})} + \log(\Delta t) & \text{if } o_{t_n} \in \{\mathtt{o1}, \mathtt{1o}\} \end{cases} \tag{5.26}$$

and,

$$F^1\left(t_n, a, D\right) \coloneqq \begin{cases} \hat{h}_{t_{n-1}}^{\Delta t}(a-1) \cdot \exp(D) - \hat{h}_{t_{n-1}}^{\Delta t}(a) & \text{if } o_{t_n} = \mathtt{oo} \\ \log \lambda_{\text{high}} + \hat{h}_{t_{n-1}}^{\Delta t}(a-1) \cdot \exp(D) - \hat{h}_{t_{n-1}}^{\Delta t}(a) & \text{if } o_{t_n} = \mathtt{o1} \\ \log \lambda_{\text{low}} + \hat{h}_{t_{n-1}}^{\Delta t}(a-1) \cdot \exp(D) - \hat{h}_{t_{n-1}}^{\Delta t}(a) & \text{if } o_{t_n} = \mathtt{1o} \end{cases} \tag{5.27}$$

The function F^0 only differs from F^1 by swapping the places of λ_{low} and λ_{high} in equation (5.27). We are now able to rewrite (5.14) as follows:

$$\Delta \mathrm{x}_{i,a}(t_n) = w(t_n) + F^i\left(t_n, a, \mathrm{x}_{1-i,a-1}(t_{n-1}) - \mathrm{x}_{i,a}(t_{n-1})\right) \tag{5.28}$$

Our strategy is to take the continuum limit of an auxiliary process, $\mathrm{y}_{i,a}(t_n)$, with simpler update equations,

$$\Delta \mathrm{y}_{i,a}(t_n) = F^i\left(t_n, a, \mathrm{y}_{1-i,a-1}(t_{n-1}) - \mathrm{y}_{i,a}(t_{n-1})\right). \tag{5.29}$$

We now prove how knowledge of the system (5.29) is sufficient for recovering the continuum limit of $\mathrm{x}_{i,a}(t_n)$ in equation (5.28).

Recovering original system

We prove three points in this section, regarding the processes $\mathrm{x}_{i,a}(t_n)$, $\mathrm{y}_{i,a}(t_n)$, defined above, and their respective continuum limits. Note that our results hold as long as the initial conditions for both systems are identical, that is, $\mathrm{x}_{i,a}(0) = \mathrm{y}_{i,a}(0)$, for all $i \in \Lambda, a < A$.

Claim 1. *For all $a_1, a_2 \in \mathbb{N}$, for all $s_1, s_2 \in \Lambda$, and for all t_n in the partition points,*

$$\mathrm{x}_{s_1,a_1}(t_n) - \mathrm{x}_{s_2,a_2}(t_n) = \mathrm{y}_{s_1,a_1}(t_n) - \mathrm{y}_{s_2,a_2}(t_n). \tag{5.30}$$

Proof. We first prove Eq. (5.30) for $a_1 = a - 1$ and $a_2 = a$ by induction on t_n. It is true for $t_0 = 0$ by assumption of the initial conditions. Next, if it is true for time point t_{n-1}, we have:

$$\begin{aligned}
\Delta\left(\mathrm{x}_{s_1,a-1}(t_n) - \mathrm{x}_{s_2,a}(t_n)\right) &\coloneqq \left[\mathrm{x}_{s_1,a-1}(t_n) - \mathrm{x}_{s_2,a}(t_n)\right] - \left[\mathrm{x}_{s_1,a-1}(t_{n-1}) - \mathrm{x}_{s_2,a}(t_{n-1})\right] \\
&= \Delta\mathrm{x}_{s_1,a-1}(t_n) - \Delta\mathrm{x}_{s_2,a}(t_n) \\
&= F^{s_1}\left(t_n, a-1, \mathrm{x}_{1-s_1,a-2}(t_{n-1}) - \mathrm{x}_{s_1,a-1}(t_{n-1})\right) \\
&\qquad - F^{s_2}\left(t_n, a, \mathrm{x}_{1-s_2,a-1}(t_{n-1}) - \mathrm{x}_{s_2,a}(t_{n-1})\right) \\
&= F^{s_1}\left(t_n, a-1, \mathrm{y}_{1-s_1,a-2}(t_{n-1}) - \mathrm{y}_{s_1,a-1}(t_{n-1})\right) \\
&\qquad - F^{s_2}\left(t_n, a, \mathrm{y}_{1-s_2,a-1}(t_{n-1}) - \mathrm{y}_{s_2,a}(t_{n-1})\right) \\
&= \Delta(\mathrm{y}_{s_1,a-1}(t_n) - \mathrm{y}_{s_2,a}(t_n)),
\end{aligned}$$

which entails that it is true for t_n. By induction on a, the general result follows. □

We will show in section 5.3.3 how the continuum limit of $\mathrm{y}_{i,a}(t_n)$ may be taken. Claim 1 implies both that the continuum of $\mathrm{x}_{i,a}(t_n)$ exists and that equations (5.30) hold for all continuous time point $t \in [0, T]$. The next claim shows us how the original system might be recovered from the auxiliary process via a normalization procedure.

Claim 2. *Whether t corresponds to a time point of the discrete or the continuous time system, $\mathrm{x}_{i,a}(t)$ may be recovered from $\mathrm{y}_{i,a}(t)$ via the following equations:*

$$\mathrm{x}_{i,a}(t) = \mathrm{y}_{i,a}(t) - K_t, \tag{5.31}$$

where,

$$K_t = \log \sum_a \left(\exp\left(\mathrm{y}_{1,a}(t)\right) + \exp\left(\mathrm{y}_{0,a}(t)\right)\right).$$

Proof. From results 1 and 2, we obtain $\mathrm{x}_{s,a}(t) - \mathrm{x}_{s_0,a_0}(t) = \mathrm{y}_{s,a}(t) - \mathrm{y}_{s_0,a_0}(t)$ for any fixed a_0 and s_0. Then, $\mathrm{x}_{s,a}(t) = \mathrm{y}_{s,a}(t) + \mathrm{x}_{s_0,a_0}(t) - \mathrm{y}_{s_0,a_0}(t)$ for all a and all s. So, we have that $\mathrm{x}_{s,a}(t) = \mathrm{y}_{s,a}(t) - K_t$, where $K_t = \mathrm{y}_{s_0,a_0}(t) - \mathrm{x}_{s_0,a_0}(t)$ does not depend on a nor on s.

Now, to find K_t in $\mathrm{x}_{s,a}(t) = \mathrm{y}_{s,a}(t) - K_t$ we take

$$\exp\left(\mathrm{x}_{s,a}(t)\right) = \frac{\exp\left(\mathrm{y}_{s,a}(t)\right)}{\exp\left(K_t\right)}. \tag{5.32}$$

Then, since $\exp\left(\mathrm{x}_{i,a}(t)\right)$ is encoding probabilities, we obtain

$$1 = \frac{\sum_a \left[\exp\left(\mathrm{y}_{1,a}(t)\right) + \exp\left(\mathrm{y}_{0,a}(t)\right)\right]}{\exp\left(K_t\right)}. \tag{5.33}$$

Then,

$$K_t = \log \sum_a \left(\exp\left(\mathrm{y}_{1,a}(t)\right) + \exp\left(\mathrm{y}_{0,a}(t)\right)\right).$$

□

5.3.3 Continuum limit

Let $t \in [0, T]$ be any *non*-click time, and $(t_n)_{n\in\mathbb{N}}$ the sequence of times that converges to t as $\Delta t \to 0$. Then, we have already established how equations (5.29) yield the following system of ODEs, indexed by $(i, a) \in \Lambda \times \mathbb{N}_{<A}$:

$$\frac{d\mathrm{y}_{i,a}(t)}{dt} = \frac{a+\alpha-1}{t+\beta} \exp\left(\mathrm{y}_{1-i,a-1}(t) - \mathrm{y}_{i,a}(t)\right) - \frac{a+\alpha}{t+\beta}$$

When t is a $\mathfrak{o1}$-click time, and $(t_n)_{n\in\mathbb{N}}$ is defined as before, the continuum limits become:

$$\begin{aligned}
\frac{d\mathrm{y}_{1,a}(t)}{dt} &= \log(\lambda_{\text{high}}) + \frac{a+\alpha-1}{t+\beta} \exp\left(\mathrm{y}_{0,a-1}(t) - \mathrm{y}_{1,a}(t)\right) - \frac{a+\alpha}{t+\beta}, \\
\frac{d\mathrm{y}_{0,a}(t)}{dt} &= \log(\lambda_{\text{low}}) + \frac{a+\alpha-1}{t+\beta} \exp\left(\mathrm{y}_{1,a-1}(t) - \mathrm{y}_{0,a}(t)\right) - \frac{a+\alpha}{t+\beta}
\end{aligned}$$

The $\mathfrak{10}$-click time case is similar, exchanging the places of λ_{low} and λ_{high} in the equations.

If we denote by $\{r_i\}_{i\in R}$ and $\{\ell_i\}_{i\in L}$ the times of the $\mathtt{01}$ and $\mathtt{10}$ observations respectively, we may combine the previous equations as follows:

$$\frac{d\mathrm{y}_{i,a}(t)}{dt} = \sum_{j\in R} C^i_{\mathtt{01}}\delta\left(t-r_j\right) + \sum_{k\in L} C^i_{\mathtt{10}}\delta\left(t-\ell_k\right) + \frac{a+\alpha-1}{t+\beta}\exp\left(\mathrm{y}_{1-i,a-1}(t)-\mathrm{y}_{i,a}(t)\right) - \frac{a+\alpha}{t+\beta}, \tag{5.34}$$

where $C^1_{01} = C^0_{10} = \log\lambda_{\text{high}}$ and $C^1_{10} = C^0_{01} = \log\lambda_{\text{low}}$ as given by table 5.1. We

observation	C^1	C^0
$\mathtt{01}$	$\log\lambda_{\text{high}}$	$\log\lambda_{\text{low}}$
$\mathtt{10}$	$\log\lambda_{\text{low}}$	$\log\lambda_{\text{high}}$

Table 5.1: **Definition of the constants in front of the delta-jumps in equation** (5.34)**.**

finally implement the following simplifications to the model. When t is a $\mathtt{01}$-click time, we can consider:

$$\frac{d\mathrm{y}_{1,a}(t)}{dt} = \log\left(\frac{\lambda_{\text{high}}}{\lambda_{\text{low}}}\right) + \frac{a+\alpha-1}{t+\beta}\exp\left(\mathrm{y}_{0,a-1}(t)-\mathrm{y}_{1,a}(t)\right) - \frac{a+\alpha}{t+\beta}$$
$$\frac{d\mathrm{y}_{0,a}(t)}{dt} = \frac{a+\alpha-1}{t+\beta}\exp\left(\mathrm{y}_{1,a-1}(t)-\mathrm{y}_{0,a}(t)\right) - \frac{a+\alpha}{t+\beta}$$

When t is a $\mathtt{10}$-click time, we can consider:

$$\frac{d\mathrm{y}_{1,a}(t)}{dt} = \frac{a+\alpha-1}{t+\beta}\exp\left(\mathrm{y}_{0,a-1}(t)-\mathrm{y}_{1,a}(t)\right) - \frac{a+\alpha}{t+\beta}$$
$$\frac{d\mathrm{y}_{0,a}(t)}{dt} = \log\left(\frac{\lambda_{\text{high}}}{\lambda_{\text{low}}}\right) + \frac{a+\alpha-1}{t+\beta}\exp\left(\mathrm{y}_{1,a-1}(t)-\mathrm{y}_{0,a}(t)\right) - \frac{a+\alpha}{t+\beta}$$

So, for $i\in\Lambda$, we obtain[5]:

$$\frac{d\mathrm{y}_{i,a}(t)}{dt} = \kappa\sum_{j\in R} i\delta\left(t-r_j\right) + \kappa\sum_{k\in L}(1-i)\delta\left(t-\ell_k\right) + \frac{a+\alpha-1}{t+\beta}\exp\left(\mathrm{y}_{1-i,a-1}(t)-\mathrm{y}_{i,a}(t)\right) - \frac{a+\alpha}{t+\beta}$$

[5]With the convention that $0\delta(0)=0$.

5.4 Summary

The recursive equations for the evolution of the joint posterior over the state and change-point count may be transformed, via a limiting procedure, into a system of continuous time differential equations. For continuously arriving evidence, the resulting equations are SDEs. Just as in the known hazard rate case, the interpretation of these decision variables requires some care. Their distribution approximates that of the log posterior of an ideal observer, for a fixed realization of the environment. For the case of pulsatile evidence, however, such limitations don't apply. The decision variables from the jump ODEs *are exactly*[6] the log posterior values of an ideal observer.

The systems of equations obtained in this section, albeit truncated, remain intractable as such. This calls for approximation methods. Of course, the question of what algorithm and mechanisms the animal brain uses remains open.

[6]Up to a normalization constant.

Chapter 6

Conclusions

In this work, we have developed and studied mathematical models of sequential decision making that find applications in contemporary neuroscience experiments.

One such experiment is the dynamic clicks task [34] in which rats must identify which of two inhomogeneous Poisson streams possesses the highest click rate, at the end of a trial. We derived two classes of ideal observer models for this task. The first model is one that assumes the hazard rate *known*. Although already present in [34], we hope to have established, in section 2.2.5, a solid mathematical formalism to explain why this model is Bayes-optimal. The second model is one that *learns* the hazard rate. We showed in section 5.3 how this latter model shares great similarity with a learning model for continuously arriving evidence.

Concerning the known hazard rate model, we could establish some numerical results in chapter 3. Firstly, we found that, with the correct time rescaling, only two effective task parameters determine the accuracy of the model's decisions. Secondly, we designed methods to fit, both the ideal observer model and an approximate linear model, to choice data. Our fitting method also encompasses a noisy version of each model, in which multiplicative Gaussian noise is applied to the evidence gained on each click.

Another task which falls within our scope is the triangles task [23]. Here, a human subject is presented with a star, which is randomly sampled on every trial from one of two potential spatial distributions. The task consists in identifying the source distribution. This task lends itself well to discrete time modeling. In spite of the fact that ideal observer models assuming the hazard rate known already existed, no algorithm, to our knowledge, solved the decision making task in a Bayes-optimal way

by learning the hazard rate on-line. This is exactly the solution that we developed in chapter 4.

The third task at which our models are directly targeted is the dots reversal task [23]. On each trial, a cloud of moving dots appears, a fraction of which moves coherently in one of two directions. This direction of motion alternates back and forth during the trial at unpredictable times, while the remaining dots move in random directions. The task is to identify the direction in which the coherent dots were moving at the end of the trial. For this task, a continuous time Bayes-optimal model may be derived as a continuum limit of the discrete time model from the triangles task. The case of a known hazard rate was published in [23] and [49]. In chapter 5, we extended these models to the case of an unknown hazard rate.

An overarching question of theoretical neuroscience is: *What computations does the brain perform?* One may also ask whether some algorithmic principles exist, to help understand such computations. A fruitful hypothesis is that *Bayesian inference* is a candidate such principle [25, 7]. At the same time, there are countless ways in which such hypothesis may be challenged [39]. The main reason is that it is extremely hard to probe what representations an animal brain possesses of a likelihood function, a prior, and a reward structure. Aside from these almost neurophysiological questions, we believe that sutdying Bayesian models of sequential decision making is useful for at least three reasons.

First, even considering the animal brain as a black box, the experimenter still has the freedom to impose the mathematical structure of their choice on the task. Bayesian inference can be seen as a mathematical exploitation of such structure. As scientists, deriving and analyzing Bayes-optimal models provides us, at the very least, with an intuition on what task components are relevant to maximizing the reward. Our determination of the effective parameters in the dynamic clicks task is a good example of this. Second, ideal observer models allow a quantitative benchmarking of accuracy and other dependent task variables. For each combination of task parameters, they provide an upper bound on the animal performance that we can expect. Last but not least, Bayes-optimal models are often a stepping stone towards approximation schemes that may be more realistic for the brain to implement [27, 48].

Appendix A

Numerical methods for free response protocol

The free response protocol is simulated by evolving the update equation (4.19) and subsequently computing the log likelihood ratio $y_n := \log R_n$ using equation. (4.20) at each timestep n. Each point along the curves in figure 4.4C corresponds to an average waiting time and average performance corresponding to a threshold value θ over 100,000 simulations. For each value of θ, the simulation is terminated when $|y_n| > \theta$ and the choice is given by the sign of y_n. To avoid excessively long simulations, we removed any that lasted longer than $n = 5000$, but we found changing this upper bound did not affect averages considerably. There were 400 values of θ chosen, discretizing the interval from $\theta = 0$ to $\theta = 3.89$.

Appendix B

Misattribution noise for pulsatile evidence

We now repeat the derivations from section 2.2.5, with the relaxed assumption that the location of each click may be miss-attributed with fixed probability q. We do not model the possibilities of a 'Miss' nor of a 'False Alarm' for perceiving a click. We start by introducing a new piece of notation:

- Given a true stimulus o_{true} at a single time point, we denote by o the *perceived* stimulus.
- Hence, $q := \Pr(o \neq o_{\text{true}})$ for $o, o_{\text{true}} \in \{\texttt{01}, \texttt{10}\}$, and $\Pr(o \neq o_{\text{true}}) = 0$ otherwise.
- The likelihood for the perceived stimulus is written $\tilde{f}$, as opposed to f in section 2.2.5.

The full likelihoods in a small interval Δt are:

$$\begin{aligned}
\tilde{f}_1^{\Delta t}(\texttt{01}) &= \Pr\left(o_{\text{true}} = \texttt{01}|+\right) \Pr\left(o = \texttt{01}|o_{\text{true}} = \texttt{01}, +\right) \\
&\quad + \Pr\left(o_{\text{true}} = \texttt{10}|+\right) \Pr\left(o = \texttt{01}|o_{\text{true}} = \texttt{10}, +\right) && \text{(B.1)} \\
&= q f_1^{\Delta t}(\texttt{10}) + (1-q) f_1^{\Delta t}(\texttt{01}) && \text{(B.2)} \\
&= q(\lambda_{\text{low}}\Delta t + o(\Delta t)) + (1-q)(\lambda_{\text{high}}\Delta t + o(\Delta t)) && \text{(B.3)} \\
&= \left[q\lambda_{\text{low}} + (1-q)\lambda_{\text{high}}\right] \cdot \Delta t + o(\Delta t) && \text{(B.4)}
\end{aligned}$$

And similarly,

$$\tilde{f}_1^{\Delta t}(\texttt{10}) = \left[q\lambda_{\text{high}} + (1-q)\lambda_{\text{low}}\right] \cdot \Delta t + o(\Delta t) \tag{B.5}$$

$$\tilde{f}_0^{\Delta t}(\texttt{01}) = \left[q\lambda_{\text{high}} + (1-q)\lambda_{\text{low}}\right] \cdot \Delta t + o(\Delta t) \tag{B.6}$$

$$\tilde{f}_0^{\Delta t}(\texttt{10}) = \left[q\lambda_{\text{low}} + (1-q)\lambda_{\text{high}}\right] \cdot \Delta t + o(\Delta t) \tag{B.7}$$

In all other cases, $\tilde{f}^{\Delta t} = f^{\Delta t}$ and equations (2.19) still hold. Equation (2.7) now becomes:

$$\log \frac{\tilde{f}_1^{\Delta t}(o)}{\tilde{f}_0^{\Delta t}(o)} = \begin{cases} \log \dfrac{\left[q\lambda_{\text{high}} + (1-q)\lambda_{\text{low}}\right] \cdot \Delta t + o(\Delta t)}{\left[q\lambda_{\text{low}} + (1-q)\lambda_{\text{high}}\right] \cdot \Delta t + o(\Delta t)}, & \text{if } o = \texttt{10} \\ \log \dfrac{\left[q\lambda_{\text{low}} + (1-q)\lambda_{\text{high}}\right] \cdot \Delta t + o(\Delta t)}{\left[q\lambda_{\text{high}} + (1-q)\lambda_{\text{low}}\right] \cdot \Delta t + o(\Delta t)}, & \text{if } o = \texttt{01} \\ \log \dfrac{o(\Delta t)}{o(\Delta t)}, & \text{if } o = 11 \\ \log \dfrac{1 - (\lambda_{\text{low}} + \lambda_{\text{high}})\Delta t + o(\Delta t)}{1 - (\lambda_{\text{low}} + \lambda_{\text{high}})\Delta t + o(\Delta t)}, & \text{if } o = \texttt{00} \end{cases} \tag{B.8}$$

The SPRT still holds with the new likelihoods. Equation (2.20) becomes:

$$\log \frac{\tilde{f}^+(o)}{\tilde{f}^-(o)} = \begin{cases} \log \dfrac{q\lambda_{\text{high}} + (1-q)\lambda_{\text{low}}}{q\lambda_{\text{low}} + (1-q)\lambda_{\text{high}}} = -\kappa_q, & \text{if } o = \texttt{10} \\ \log \dfrac{q\lambda_{\text{low}} + (1-q)\lambda_{\text{high}}}{q\lambda_{\text{high}} + (1-q)\lambda_{\text{low}}} =: \kappa_q, & \text{if } o = \texttt{01} \\ 0, & \text{if } o = 11 \\ 0, & \text{if } o = \texttt{00} \end{cases} \tag{B.9}$$

Since $q \in [0,1]$, the numerator and denominator in $\tilde{f}^+(o)/\tilde{f}^-(o)$ are convex linear combinations of λ_{low} and λ_{high}. We deduce the following: $|\kappa_q| \leq |\kappa|$ and:

- $\kappa \cdot \kappa_q > 0$ if $q < 1/2$
- $\kappa \cdot \kappa_q < 0$ if $q > 1/2$
- $\kappa_q = 0$ if $q = 1/2$
- $\kappa_q = \kappa$ if $q = 0$ and $\kappa_q = -\kappa$ if $q = 1$

In this setting, the new proxy for the SNR would be κ_q, and we can observe that for $0 \leq q \leq 1/2$, κ_q is a decreasing function of q. Values of $q > 1/2$, when q is known to the ideal-observer, is equivalent to inverting the favored states of the right- and left-perceived clicks.

Bibliography

[1] Brodylab's official web page. http://brodylab.org/ (Last Accessed 07/07/2018).

[2] Gold Lab's official web page. http://www.med.upenn.edu/goldlab/# (Last Accessed 07/07/2018).

[3] Video of the dots reversal stimulus. https://doi.org/10.7554/eLife.08825.012 (Last Accessed 07/07/2018).

[4] Ryan P. Adams and David J. C. MacKay. Bayesian online changepoint detection. *ArXiv e-prints*, 2007.

[5] Gábor Balázsi, Alexander van Oudenaarden, and James J. Collins. Cellular decision making and biological noise: from microbes to mammals. *Cell*, 144(6):910–925, 2011.

[6] Leonard E. Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Statistics*, 41(1):164–171, 1970.

[7] Jeffrey M. Beck, Wei J. Ma, Roozbeh Kiani, Timothy D. Hanks, Anne K. Churchland, Jamie D. Roitman, Michael N. Shadlen, Peter E. Latham, and Alexandre Pouget. Probabilistic population codes for Bayesian decision making. *Neuron*, 60(6):1142–1152, 2008.

[8] Sebastian Bitzer, Hame Park, Felix Blankenburg, and Stefan J. Kiebel. Perceptual decision making: drift-diffusion model is equivalent to a Bayesian model. *Frontiers in Human Neuroscience*, 8(102):1–17, 2014.

[9] Isabelle Blanchette and Anne Richards. The influence of affect on higher level cognition: a review of research on interpretation, judgement, decision making and reasoning. *Cognition & Emotion*, 24(4):561–595, 2010.

[10] Rafal Bogacz, Eric Brown, Jeff Moehlis, Philip Holmes, and Jonathan D. Cohen. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, 113(4):700–765, 2006.

[11] Kenneth H. Britten, Michael N. Shadlen, William T. Newsome, and J. Anthony Movshon. Responses of neurons in macaque MT to stochastic motion signals. *Visual Neuroscience*, 10(6):1157–1169, 1993.

[12] Bingni W. Brunton, Matthew M. Botvinick, and Carlos D. Brody. Rats and humans can optimally accumulate evidence for decision-making. *Science*, 340(6128):95–98, 2013.

[13] Olivier Cappé, Eric Moulines, and Tobias Rydén. *Inference in Hidden Markov Models*. Springer Series in Statistics. Springer, New York, NY, 2005.

[14] Christopher P. Chambers, Federico Echenique, and Kota Saito. Testing theories of financial decision making. *Proceedings of the National Academy of Sciences*, 113(15):4003–4008, 2016.

[15] Gabriela Czanner, Sridevi V. Sarma, Demba Ba, Uri T. Eden, Wei Wu, Emad Eskandar, Hubert H. Lim, Simona Temereanca, Wendy A. Suzuki, and Emery N. Brown. Measuring the signal-to-noise ratio of a neuron. *Proceedings of the National Academy of Sciences*, 112(23):7141–7146, 2015.

[16] Arthur P. Dempster, Nan M. Laird, and Donald B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[17] Sophie Deneve. Bayesian spiking neurons I : inference. *Neural Computation*, 20(1):91–117, 2008.

[18] Ben Deverett, Sue A. Koay, Marlies Oostland, and Samuel S.-H. Wang. A cerebellar role in evidence-guided decision-making. *bioRxiv*, 2018.

[19] James P. Egan. *Signal Detection Theory and ROC analysis*. Academic Press Series in Cognition and Perception. London, UK: Academic Press, 1975.

[20] Wim C. V. Etten. *Introduction to random signals and noise*. John Wiley & Sons, 2005.

[21] Crispin W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*. Springer complexity. Springer, 2004.

[22] Christopher M. Glaze, Alexandre L. S. Filipowicz, Joseph W. Kable, Vijay Balasubramanian, and Joshua I. Gold. A bias-variance trade-off governs individual differences in on-line learning in an unpredictable environment. *Nature Human Behaviour*, 2(3):213–224, 2018.

[23] Christopher M. Glaze, Joseph W. Kable, and Joshua I. Gold. Normative evidence accumulation in unpredictable environments. *eLife*, 4:e08825, 2015.

[24] Joshua I. Gold and Michael N. Shadlen. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, 36(2):299–308, 2002.

[25] Joshua I. Gold and Michael N. Shadlen. The neural basis of decision making. *Annual Review of Neuroscience*, 30:535–574, 2007.

[26] David M. Green and John A. Swets. *Signal detection theory and psychophysics*. Wiley: New York, New York, 1966.

[27] Anna Kutschireiter, Simone C. Surace, Henning Sprekeler, and Jean-Pascal Pfister. Nonlinear Bayesian filtering and learning: a neuronal dynamics for perception. *Scientific Reports*, 7:8722, 2017.

[28] Olivier Lebeltel, Pierre Bessiere, Julien Diard, and Emmanuel Mazer. Bayesian robot programming. *Autonomous Robots*, 16(1):49–79, 2004.

[29] Tyler McMillen and Philip Holmes. The dynamics of choice among multiple alternatives. *Journal of Mathematical Psychology*, 50(1):30–57, 2006.

[30] Peter Mörters and Yuval Peres. *Brownian motion*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2010.

[31] Jerzy Neyman and Egon S. Pearson. IX. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 231:289–337, 1933.

[32] Khanh P. Nguyen, Krešimir Josić, and Zachary P. Kilpatrick. Optimizing sequential decisions in the drift-diffusion model. *ArXiv e-prints*, 2018.

[33] Onyekachi Odoemene, Sashank Pisupati, Hien Nguyen, and Anne K. Churchland. Visual evidence accumulation guides decision-making in unrestrained mice. *bioRxiv*, 2017.

[34] Alex T. Piet, Ahmed E. Hady, and Carlos D. Brody. Rats optimally accumulate and discount evidence in a dynamic environment. *bioRxiv*, 2017.

[35] Lucas Pinto, Sue A. Koay, Ben Engelhard, Alice M. Yoon, Ben Deverett, Stephan Y. Thiberge, Ilana Witten, David W. Tank, and Carlos D. Brody. An accumulation-of-evidence task using visual pulses for mice navigating in virtual reality. *bioRxiv*, 2018.

[36] Did (pseudonym). How exactly are the beta and gamma distributions related? Mathematics Stack Exchange, https://math.stackexchange.com/q/190841 (Last Accessed 07/07/2018).

[37] Lawrence Rabiner and Biing-Hwang Juang. An introduction to Hidden Markov Models. *IEEE ASSP Magazine*, 3:4–16, 1986.

[38] Adrian E. Radillo, Alan Veliz-Cuba, Zachary P. Kilpatrick, and Krešimir Josić. Evidence accumulation and change rate inference in dynamic environments. *Neural Computation*, 29(6):1561–1610, 2017.

[39] Dobromir Rahnev and Rachel Denison. Suboptimality in perception. *Behavioral and Brain Sciences*, pages 1–107, 2018.

[40] Roger Ratcliff, Philip L. Smith, Scott D. Brown, and Gail McKoon. Diffusion decision model: current issues and history. *Trends in Cognitive Sciences*, 20(4):260–281, 2016.

[41] Joseph F. Ritt. On the zeros of exponential polynomials. *Transactions of the American Mathematical Society*, 31:680–686, 1929.

[42] Jamie D. Roitman and Michael N. Shadlen. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *The Journal of Neuroscience*, 22(21):9475–9489, 2002.

[43] Barbara Sakitt. Indices of discriminability. *Nature*, 241(5385):133–134, 1973.

[44] C. Daniel Salzman, Kenneth H. Britten, and William T. Newsome. Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, 346(6280):174–177, 1990.

[45] Thomas D. Seeley, P. Kirk Visscher, Thomas Schlegel, Patrick M. Hogan, Nigel R. Franks, and James A. R. Marshall. Stop signals provide cross inhibition in collective decision-making by honeybee swarms. *Science*, 335(6064):108–111, 2012.

[46] Michael N. Shadlen and William T. Newsome. Motion perception: seeing and deciding. *Proceedings of the National Academy of Sciences*, 93(2):628–633, 1996.

[47] John G. Skellam. The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society*, 109(3):296, 1946.

[48] Sacha Sokoloski. Implementing a Bayes filter in a neural circuit: the case of unknown stimulus dynamics. *Neural Computation*, 29(9):2450–2490, 2017.

[49] Alan Veliz-Cuba, Zachary P. Kilpatrick, and Krešimir Josić. Stochastic models of evidence accumulation in changing environments. *Society for Industrial and Applied Mathematics Review*, 58(2):264–289, 2016.

[50] Abraham Wald. *Sequential analysis*. Wiley series in Probability and Mathematical Statistics. John Wiley & Sons, 1947.

[51] Robert C. Wilson, Matthew R. Nassar, and Joshua I. Gold. Bayesian online learning of the hazard rate in change-point problems. *Neural Computation*, 22(9):2452–2476, 2010.