Modelling a simple choice task: Stochastic dynamics of mutually inhibitory neural groups

Eric Brown[1] and Philip Holmes[1,2]

[1] Program in Applied and Computational Mathematics,
Princeton University, Princeton, NJ 08544, U.S.A.
[2] Department of Mechanical and Aerospace Engineering,
Princeton University, Princeton, NJ 08544, U.S.A.

(Appears in *Stochastics and Dynamics* **1**:2 (2001), 159–191)

Abstract

We describe the dynamical and bifurcational behavior of two mutually inhibitory, leaky, neural units subject to external stimulus, random noise, and 'priming biases.' The model describes a simple forced choice experiment and accounts for varying levels of expectation and control. By projecting the model's dynamics onto slow manifolds, using judicious linear approximations, and solving for one-dimensional (reduced) probability densities, analytical estimates are developed for reaction time distributions and shown to compare satisfactorily with 'full' numerical data. A sensitivity analysis is performed and the effects of parameters assessed. The predictions are also compared with behavioral data. These results may help correlate low-dimensional models of stochastic neural networks with cognitive test data, and hence assist in parameter choices and model building.

1 Introduction and motivation

In this paper we examine the dynamics of two inhibitory neural units. Our motivation is the correlation of low-dimensional neural models with behavioral observables of cognition, such as such reaction time (RT) and error rate (ER) in simple decision tasks. We assume that there are dedicated populations of neurons, modeled as parallel distributed processing (PDP) units [23], that are selectively responsive to different stimuli. Each unit accumulates 'activation' (a population-averaged analogy to membrane voltage) in

response to external stimuli, loses activation through a decay term, and may gain and lose it due to connections with itself and with the other units in the network. The units are also subjected to additive noise.

Our connectionist [1, 23] model is based on that of Usher and McClelland [28], and builds on their analysis. We adopt a logistic activation function and modifications due to Botvinick [6] and R. Cho [7] that add control parameters modelling the effects of conflict and expectation. There are ten parameters in the trial dynamics model, three of which (biases) are in turn updated by rules based on previous trial outcomes, involving three time constants and three reference levels. The resulting $10-3+6 = 13$-parameter family of dynamical systems displays a rich behavioral repertoire which is difficult to characterize by simulation. Here we perform analyses to assist in parameter choices and in understanding the resulting dynamics. A related (noise-free) 'multivibrator' system is discussed in [2].

The paper proceeds as follows. Section 2 introduces the stochastic ODEs describing the network model, which are analyzed in Section 3. Section 3.1 provides a preliminary analysis of the noise-free problem, including bifurcation behavior. Section 3.2 introduces the Komolgorov (Fokker-Planck) formulation of the problem, and Sections 3.3-3.5 develop numerical and analytical methods for its solution. In particular, we derive closed form approximate expressions for statistics of reaction time distributions and demonstrate their general validity. Section 4 contains a discussion of the effects of parameters in determining predicted reaction time statistics, and comparisons of our analytical predictions to experimental data. We draw conclusions in Section 5, and (as in [28]) in doing so relate the model to diffusion models from the psychological literature [21, 22].

2 Description of the model

The model involves two mutually-inhibiting, leaky, neural units characterized by state variables x_j, subject to external stimuli ρ_j (normalized so that, when $\rho_j \neq 0$, $\rho_2 = 1 - \rho_1$), additive noises modeled by independent, scaled Wiener processes $\sigma W_{t,j}$, and 'priming biases' $i_0 + b_j$, including an overall level i_0 and separate unit biases b_j. Each unit inhibits the other via an activation function $f(x; g, b) = 1/[1 + \exp(-g(x - b))]$ with gain g that achieves half level at $x = b$. (We generally omit explicit reference to g, b below and simply write $f(x)$.) The equations are:

$$dx_1 \quad = \quad (-kx_1 - \beta f(x_2) + i_0 + b_1 + \rho_1)\, dt + \sigma dW_{t,1}$$

$$
\begin{aligned}
& \triangleq && X_1(x_1, x_2)dt + \sigma dW_{t,1} \ , \\
dx_2 \quad &= && \left(-kx_2 - \beta f(x_1) + i_0 + b_2 + \rho_2\right) dt + \sigma dW_{t,2} \\
& \triangleq && X_2(x_1, x_2)dt + \sigma dW_{t,2} \ ,
\end{aligned} \tag{1}
$$

where k denotes the leak (time constant $1/k$) and β the inhibition level. These are examples of Ito diffusion equations. Following each trial the units are allowed to respond to the bias and noise inputs for a *preparatory phase* or response-to-stimulus interval (RSI) of duration τ_{P}, with $\rho_j \equiv 0$, thus establishing the initial state for the next trial *trial period* during which they integrate the inputs including the stimuli $\rho_2 = 1 - \rho_1$.

Two distinct classes of cognitive choice task may be modelled by Eqn. (1): (i) the free-response protocol, in which subjects respond to stimulus presentations as soon as a decision has been reached, and (ii), the forced-response protocol, in which they are required to respond at a fixed time t_r following stimulus presentation. In the former, a decision is declared when one of the unit activations $f(x_j)$ crosses a preset threshold θ; the instant at which this crossing occurs is the reaction time (RT) for the trial. Under the forced-response protocol any threshold crossings while $t < t_r$ are ignored, and the relative values of the activations $x_j(t_r)$ determine the decision at interrogation time t_r. In both protocols, x_j may continue to evolve and be monitored for a period following the decision itself. After this period, which in the free-response protocol may depend on the time taken for the winning unit to reach threshold, the x_j are either reset to $x_j = 0$ or otherwise relaxed prior to initiation of the following preparatory phase.

Modelling the forced-response protocol requires (Monte Carlo) solution of Eqns. (1), or, equivalently, of the corresponding Komolgorov equation, leading to expressions for the probability that trajectories have crossed threshold, and, in turn, for time-dependent probability fluxes. In contrast, reaction times under the free-response protocol are described by first passage or hitting time distributions of the stochastic processes x_j. Such distributions can be found via boundary value techniques in the backward Komolgorov formulation [20, 13]. However, we will see in Section 3.4.3 that in many cases probability fluxes also provide good approximations for hitting time densities. Thus, the forward Komolgorov formulation is useful for both protocols.

Between each trial and the following preparatory period the overall bias is updated according to a discrete rule with fading memory, characterized by a time constant $1/\lambda_{i_0}$:

$$
i_0(n+1) = \lambda_{i_0} i_0(n) + (1 - \lambda_{i_0})[i_{\max} - aC(n)] \ , \tag{2}
$$

where

$$C(n) = \int_{t(n)}^{t(n)+\tau_{\rm T}} f(x_1(t))f(x_2(t))\,dt \tag{3}$$

denotes the conflict experienced on the n'th trial, estimated as the integrated product of the unit activations. Here, $\tau_{\rm T}$ is the duration of the trial phase. Following [6], conflict is taken proportional to the product of unit activations and, via Eqn. (2), i_0 decreases (resp. increases) following episodes of high (resp. low) conflict. In Eqn. (3), conflict is integrated only over the trial itself; preparatory and post trial periods could also be included, the latter reflecting the fact that processing continues after choices have been made. Note that conflict is not normalized by trial length or preparatory periods, but rather increases monotonically with $\tau_{\rm T}$.

Individual biases may be held at $b_j = 0$, or updated depending upon prior stimuli. Such history effects have been documented in eg. [5, 25]. History-based updates depend on two factors. The first concerns the *number of neural units* which detect patterns: here, combinations of simple repetitions or alternations in unambiguous stimuli (a stimulus ρ_j will be assumed to be unambiguous, or salient, if it differs from the completely ambiguous value of $1/2$ by at least some margin $\phi > 0$). We consider the different implications of: (i) assuming there to be a total of *four* pattern detection units, one to detect repetitions and one to detect alternations for *each* of the decision units $j = 1, 2$; and (ii) assuming only *two* repetition and alternation detectors, shared between the decision units.

In the first case, the independent units determine biases through the terms b_j^A and b_j^R as follows:

$$\begin{aligned} b_j(n+1) &= \alpha_A b_j^A(n+1) + \alpha_R b_j^R(n+1), \quad \text{where} \\ b_j^A(n+1) &= \min\left[\lambda_b b_j^A(n) + (1-\lambda_b) f_j^A(n), b_{max}\right] \\ b_j^R(n+1) &= \min\left[\lambda_b b_j^R(n) + (1-\lambda_b) f_j^R(n), b_{max}\right], \end{aligned} \tag{4}$$

and α_A, α_R are weights accorded to alternation and repetition respectively. In the second case the shared units determine $b_j(n+1)$ via:

$$\begin{aligned} b_j(n+1) &= \left\{ \begin{array}{ll} \alpha_A b^A(n+1) & \text{if } \rho_j(n) < 1/2 - \phi \\ \alpha_R b^R(n+1) & \text{if } \rho_j(n) > 1/2 + \phi \\ 0 & \text{otherwise} \end{array} \right\}, \quad \text{where} \\ b^A(n+1) &= \min\left[\lambda_b b^A(n) + (1-\lambda_b) f^A(n), b_{max}\right]; \\ b^R(n+1) &= \min\left[\lambda_b b^R(n) + (1-\lambda_b) f^R(n), b_{max}\right]. \end{aligned} \tag{5}$$

Here the functions $f_j^A(n)$ and $f_j^R(n)$ in Eqn. (4) (resp. $f^A(n)$ and $f^R(n)$ in Eqn. (5)) are pattern detectors for alternating and repeating stimuli respectively, which take the values 0 or 1 for each trial. λ_b represents the decay of influence of prior stimuli.

The second choice concerns the *definitions of* $f_j^{A,R}$. Detectors may examine two previous timesteps to determine whether the most recent stimulus represents an alternation, a repetition, or neither. Then the four- and two-unit rules are, respectively:

$$f_j^A(n) = \begin{cases} 1 & \text{if } \rho_j(n-1) > 1/2+\phi \text{ and } \rho_j(n) < 1/2-\phi \\ 0 & \text{otherwise ,} \end{cases} \tag{6}$$

$$f_j^R(n) = \begin{cases} 1 & \text{if } \rho_j(n-1) > 1/2+\phi \text{ and } \rho_j(n) > 1/2+\phi \\ 0 & \text{otherwise ,} \end{cases} \tag{7}$$

and

$$f^A(n) = \begin{cases} 1 & \text{if } \rho_j(n-1) > 1/2+\phi \text{ and } \rho_j(n) < 1/2-\phi \text{ ,} \\ & j = 1 \text{ or } 2 \text{ ,} \\ 0 & \text{otherwise ,} \end{cases} \tag{8}$$

$$f^R(n) = \begin{cases} 1 & \text{if } \rho_j(n-1) > 1/2+\phi \text{ and } \rho_j(n) > 1/2+\phi \text{ ,} \\ & j = 1 \text{ or } 2 \text{ ,} \\ 0 & \text{otherwise .} \end{cases} \tag{9}$$

Alternatively, the functions $f_j^A(n)$, $f_j^R(n)$, $f^A(n)$, and/or $f^R(n)$ may 'autoprime' the biases toward repetition or alternation based on the most recent stimulus *alone*, without regard to earlier history. In this case we have

$$f_j^A(n) = \begin{cases} 1 & \text{if } \rho_j(n) < 1/2-\phi \\ 0 & \text{otherwise ,} \end{cases} \tag{10}$$

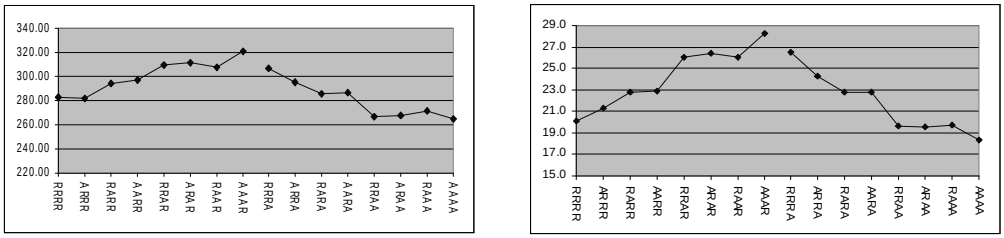
$$f_j^R(n) = \begin{cases} 1 & \text{if } \rho_j(n) > 1/2+\phi \\ 0 & \text{otherwise ,} \end{cases} \tag{11}$$

and

$$f^A(n) = \begin{cases} 1 & \text{if } \rho_j(n) > 1/2+\phi \text{ or } \rho_j(n) < 1/2-\phi, \\ & j = 1 \text{ or } 2 \\ 0 & \text{otherwise ,} \end{cases} \tag{12}$$

$$f^R(n) \equiv f^A(n) \ . \tag{13}$$

The full range of possibilities for bias update procedures includes all combinations of the alternatives represented by the rules grouped under



a) **b)**

Figure 1: Mean reaction times (RTs) for all sequence types (of stimulus length 5). R = same stimulus (repetition); A = different stimulus (alternation); letters are ordered from least (top) to most recent (bottom). For example, "RRAA" represents two repeats followed by two alternations (11121 or 22212). (a) Data adapted and replotted from [25] (Fig. 2) for 1 sec. response-to-stimulus intervals, ms. time units; (b) results of Monte-Carlo simulations, arbitrary units.

Eqns. (4) and (5) for bias updating, and Eqns. (6-7), (8-9) or (10-11), (12-13) for pattern detection. If different bias update rules are chosen for alternation and for repetition, Eqns. (4) and (5) must be modified accordingly.

To constrain the many possibilities, R. Cho et al. [7] compared sequences of reaction times generated by numerical integration of Eqns. (1). Since the past four elements of a stimulus history generally play the dominant role in determining the outcome of a current trial [25], we set $\lambda_b = 0.5$ (then $\lambda_b^4 = 0.0625$). The $2^4 = 16$ resulting possibilities are represented by strings of A's and R's representing alternations and repetitions, with the most recent stimulus pattern on the right or bottom (eg. AAAR stands for alternations on three previous trials and repetition on the current one). Data from Monte-Carlo simulations of Eqns. (1) subject to random salient ($\rho_j = 0.85$) stimulus sequences was analyzed to produce mean RTs for each stimulus sequence. In preliminary work, a reasonable match to experimental data from the quantitative literature [25] was obtained with independent neural units registering alternations and repetitions depending on the prior two stimuli (Eqn. (4) with Eqns. (6-7)) [7]. Fig. 1 displays these results for parameters given in Table 1 below, excepting $i_0 = 0$ (fixed), $\sigma = .152$, $g = 4$, $\rho_2 = .75$, and $\tau_{\rm P} = 6$. The forthcoming paper [7] will further examine stimulus history effects on reaction times and error rates.

In our preliminary analysis we found it useful to solve the difference equations (4-5) explicitly for each of the sixteen histories, again using the fact that for $\lambda_b \approx 1/2$ memory fades sufficiently rapidly that stimuli preceding the $(n-4)$'th have negligible effect on $b_j(n+1)$. Plotting the resulting pairs $(b_1(n+1), b_2(n+1))$ on the (x_1, x_2) phase plane reveals the relative importance of different choices on the weights accorded to repetition and alternation detection. The model may be put in approximate accordance with experiments by scaling time so that reaction times are comparable: see Section 4.3. Meanwhile, note the good qualitative agreement between the mean reaction times of Figs. 1 (a) and (b).

3 Analysis of the model

3.1 The phase plane and bifurcations

We give an outline of the system's behavior, followed by more detailed qualitative and quantitative studies (cf. [2]). To gain an initial understanding, we first consider Eqn. (1) without noise ($\sigma = 0$), in which case it is a dissipative system possessing a Liapunov function [14, 16, 17]. Thus, released from an initial (non-equilibrium) state, typical solutions approach asymptotically stable fixed points (sinks). Fixed points lie at intersections of *nullclines*, the curves on which the components' rates of change vanish:

$$x_1 = \frac{1}{k}[i_0 + b_1 + \rho_1 - \beta f(x_2)] \ , \ x_2 = \frac{1}{k}[i_0 + b_2 + \rho_2 - \beta f(x_1)] \ . \qquad (14)$$

There may be one, two, or three equilibria, depending upon parameter values. If there is one, it is stable; if three, two are stable and the other "central" one is an unstable saddle. Since the maximum slope of the (scaled) activation function $\beta f(x)$ in (14) is $\beta g/4$ (occurring at $x = b$), if $\beta g \leq 4k$ the nullclines intersect exactly once for any i_0, b_j, ρ_j; if $\beta g > 4k$ they may intersect in three points when $i_0 + b_j + \rho_j - \beta < bk < i_0 + b_j + \rho_j; j = 1, 2$. For equal bias and stimulus $\gamma = i_0 + b_j + \rho_j, \ j = 1, 2$, the bistability region is delimited by the condition that the symmetric fixed point $(\bar{x}, \bar{x})$ occur where the nullclines both have slope -1. Solving

$$\bar{x} = \frac{1}{k}[\gamma - \beta f(\bar{x})] \ \text{ and } \ -\frac{\beta}{k} f'(\bar{x}) = -1 \qquad (15)$$

simultaneously, we obtain $\gamma - k\bar{x} = \beta[1 \pm \sqrt{1 - (4k/\beta g)}]/2$. Thus, for $\beta g > 4k$, a pair of fixed points symmetric about the diagonal bifurcate from $(\bar{x}, \bar{x})$

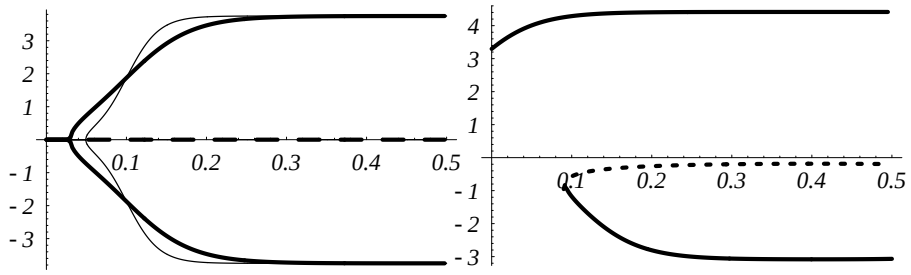


Figure 2: Bifurcation diagrams for varying i_0. Ordinate shows value of $x_2 - x_1$ to indicate symmetry breaking. (left) For $b_1 = b_2 = 0$ and the standard parameter set, a *pitchfork* bifurcation occurs from the symmetry axis $x_1 = x_2$. Results for $g = 5$ (thick lines) and $g = 10$ (thin lines); (right) for $g = 5$, $b_1 = .0011$, $b_2 = .1342$ (biases representing the AAAA stimulus history), the pitchfork bifurcation is unfolded in a "cusp catastrophe" and a *saddle-node* bifurcation occurs [14].

at

$$\bar{x}_{\rm pf} = \frac{1}{k}\left[\gamma - \frac{\beta}{2}\left(1 \pm \sqrt{1 - \frac{4k}{\beta g}}\right)\right]; \tag{16}$$

as γ varies, the bifurcations occur where

$$\gamma_{\rm pf} = k\left[b - \frac{1}{g}\ln\left(\frac{1 \pm \sqrt{1 - \frac{4k}{\beta g}}}{1 \mp \sqrt{1 - \frac{4k}{\beta g}}}\right)\right] + \frac{\beta}{2}\left(1 \mp \sqrt{1 - \frac{4k}{\beta g}}\right) \quad . \tag{17}$$

Fig. 2 shows bifurcation diagrams [14] as fixed point loci $x_2 - x_1$ in terms of i_0 for $\rho_j = 0$, $b_1 \geq b_2 = 0$ and what will be referred to as the standard parameter set (displayed in Table 1).

Fig. 3(a,b) shows nullclines for two parameter conditions, and also shows typical system states at the close of the preparatory period ($\rho_j \equiv 0$), in the presence of noise. Solutions move relatively rapidly towards the central region where the nullclines are close, and thereafter slowly drift and diffuse under the influence of noise. Increased bias i_0 shifts the (unique, stable) equilibrium of Fig. 3(a) diagonally, changing it into an unstable saddle point, and creating new stable and unstable asymmetric equilibria closer to threshold (cf. Eqn. (16) and Fig. 2). As bias continues to increase, the asymmetric equilibria recombine with the symmetric one and monostability returns. 'Noisy' solutions therefore spread towards these states, leading in Fig. 3(b)

β	.75
k	.2
σ	.158
ρ_2	.85
b	.5
g	5
i_0	.1583
τ_{P}	1
θ	.9
$b_1(AAAA)$	.0011
$b_2(AAAA)$	.1342
$b_1(AAAR)$	.1342
$b_2(AAAR)$	.0011
$b_1(eq)$	(.1342+.0011)/2
$b_2(eq)$	(.1342+.0011)/2

Table 1: The standard parameter set used to demonstrate many of the methods used in this paper. These parameters were selected from preliminary trial and error efforts to match experimental sequence history data (cf. Fig. 1) using Monte Carlo simulations of Eqns. (1) with piecewise linear activation functions. The values of $b_j(AAAA)$ and i_0 were taken from trial-to-trial averages performed using the rest of the standard parameter set, with the former averaged over only those subsequences of trials with the relevant stimulus history and i_0 averaged over all trials. These biases were used in a natural way to determine b_j for the other two stimulus histories.

to a more diffuse sample of initial data for the trial itself. During trial, with $\rho_j \neq 0$, solutions drift towards a 'new' stable equilibrium. If the stimulus is unambiguous (ρ_1 exceeds ρ_2, or *vice versa* by a sufficient margin), this is unique; if not ($\rho_1 \approx \rho_2$), bistability may persist.

The analysis above reveals that parameters should not be varied arbitrarily: the net effect of biases i_0, b_j and stimuli ρ_j is to shift equilibria rightward and upward, and hence thresholds must be set with these values in mind, so that typical distributions during the preparatory period lie below threshold, and trial equilibria lie beyond. Hence, reasonable threshold values in x_j scale linearly with i_0, b_j and ρ_j. Threshold values themselves, being computed by inverting the function $f(x_\theta) = \theta$ to give $x_\theta = b + [\ln(\theta) - \ln(1-\theta)]/g$, are most significantly affected by activation gain and bias g, b. Replacing

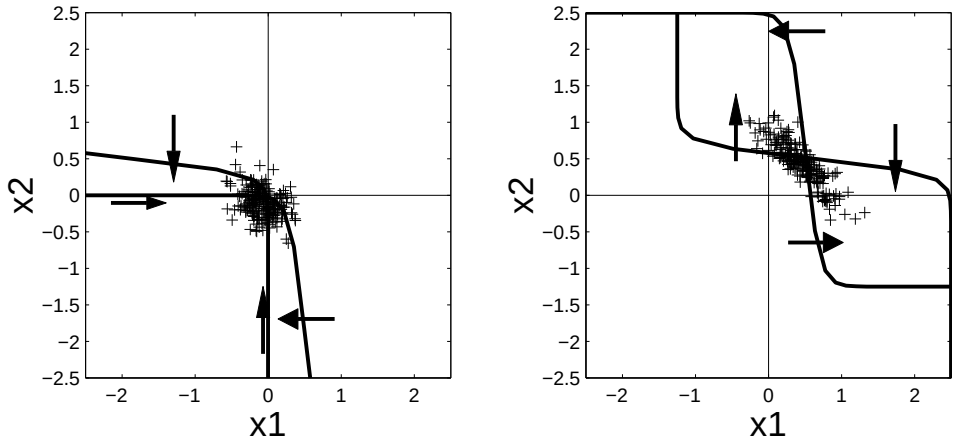


Figure 3: Nullclines and typical distributions of solutions at the end of the preparatory period, with $b_j = 0$, $\tau_{\rm P} = 2$, and other parameters other than i_0 drawn from the standard parameter set. (left) $i_0 = 0$ (following high conflict); (right) $i_0 = 0.5$ (following low conflict). Results are from Monte-Carlo numerical solution of the stochastic differential equation (1).

x_j by $x_j - b$ simply adds the term kb to each right hand side of Eqn. (1); thus b effectively specifies an 'origin' for the state space and we may keep it constant without loss of generality (here we take $b = 0.5$). Increases in g cause decreases in x_θ, and (more modest) increases in distance between trial equilibria and the diagonal $x_1 = x_2$: cf. Fig. 2.

The essential picture is that gain variations primarily affect the distance solutions must travel from stimulus onset to cross threshold, hence increases in g reduce reaction time (RT) means and vice versa. In contrast, increases in conflict i_0 and uniform increases in biases $b_{1,2}$ near the bifurcation point in Eqn. (17) cause both a shift towards threshold *and* spread of solution distributions prior to stimulus onset, and thus reduce RT means but *increase* RT variances. The extent of this spread depends on the time $\tau_{\rm P}$ allotted to the preparatory cycle. We now describe how these effects can be quantified.

3.2 Probability densities

Eqn. (1) is an example of an Ito diffusion. Let $p(x_1, x_2, t; p_0)$ denote the transition probability density at time t from initial density p_0; then p obeys

the forward Fokker-Planck or Komolgorov equation [3]:

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x_1}(X_1 p) - \frac{\partial}{\partial x_2}(X_2 p) + \frac{\sigma^2}{2}\left(\frac{\partial^2 p}{\partial x_1^2} + \frac{\partial^2 p}{\partial x_2^2}\right) , \tag{18}$$

where $X_j(x_1, x_2)$ denote the deterministic vectorfield and σ scales the (independent) Wiener processes in Eqn. (1). During a trial p evolves, first in response to the preparatory vectorfield ($\rho_j \equiv 0$) and then to the stimuli. If p can be computed or approximated, we may integrate the probability mass which has crossed threshold at any given time and hence find the probability that a particular decision will be made at a particular time t in the forced-response protocol. We will see in Section 3.4.3 how this may be related to free-choice RT distributions. Below we sometimes drop explicit reference to p_0 and write $p(x_1, x_2, t)$.

3.3 Numerical solutions

The variable coefficient linear PDE (18) cannot generally be solved explicitly, but numerical solutions are routine. Fig. 4 (a,b) shows $p(x_1, x_2; t)$ at preparatory cycle end and during trial. The initial distribution for the simulation was a symmetric Gaussian with variance .04 centered at $(x_1, x_2) = (0, 0)$. Computations were done with an adaptive-grid finite element algorithm (*FlexPDE* [4]) and the standard parameter set of Table 1.

Fig. 4(c) shows the threshold crossing flux $R_2^f(t)$ of probability, which was computed as the flux $\mathbf{J} = p\mathbf{X} - \frac{\sigma^2}{2}\nabla p$ of probability numerically integrated across the relevant decision thresholds ($x_\theta, x_2 \leq x_\theta$) and ($x_1 \leq x_\theta, x_\theta$). These thresholds were used for simulation of salient stimuli, and motivated by the notion of hitting times, do not include post-threshold segments (in the forced-response protocol, these segments could be relevant for trajectories that move between post-threshold regions without crossing first into the region $x_j < x_\theta$, $j = 1, 2$). Hence, the numerical results for threshold crossing fluxes presented here are valid under the assumption that, as in the case shown in Fig. 4, this type of recrossing event does not occur with significant probability. Finally, Fig. 4(d) shows the probability $P_2^f(t)$ that the correct choice would have been made at time $t = t_r$ after the start of the trial in the forced-response protocol, obtained by numerical integration of the *FlexPDE* results over the region $x_2 \geq x_\theta$.

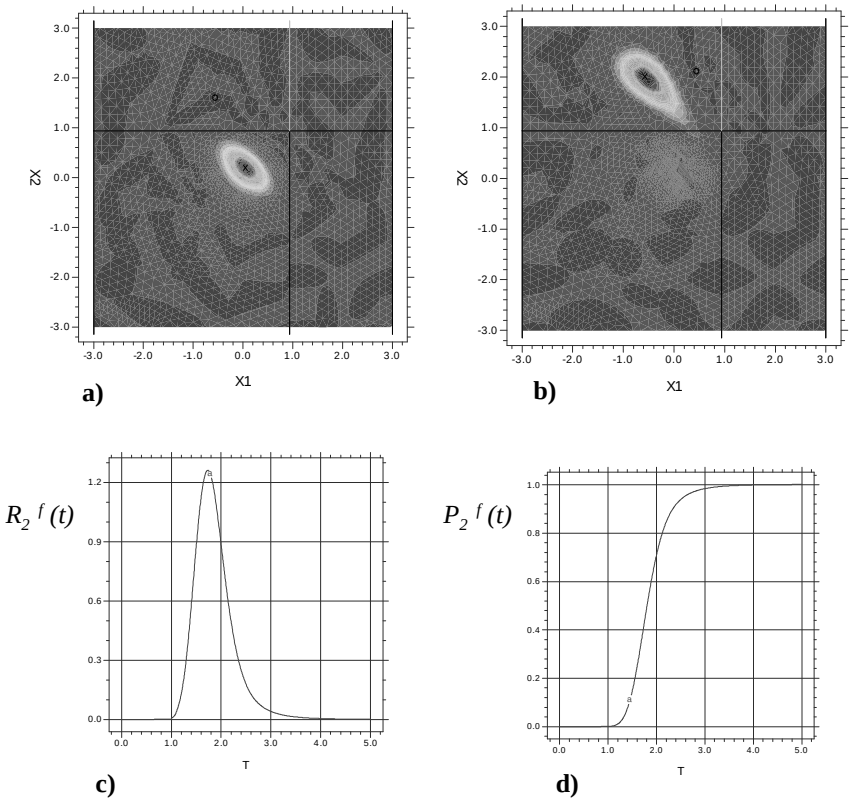


Figure 4: Computations of (a) $p(x_1, x_2, t = \tau_p; p_0 \approx \delta(0,0))$ and (b) $p(x_1, x_2, t = 3)$ during a trial with the standard parameter set for the AAAA stimulus history, computed using *FlexPDE* ; (c) flux distribution $R_2^f(t)$; (d) integral of $p(x_1, x_2, t = \tau_p)$ over region past threshold 2. Irregular backgrounds away from the main peaks in (a) and (b) are graphical artefacts due to low probability values.

3.4 Analytical approximations

3.4.1 General method

Here we outline an approximation developed from Stone and Holmes [26]; also see Chapter 8 of [15]. As Fig. 3 indicates, the two nullclines are close in the central region of phase space, and the deterministic vectorfield is directed relatively strongly, parallel to the diagonal, towards this region. We therefore *assume* that, after an initial transient, the bulk of the preparatory cycle involves evolution in response to the component of the vectorfield projected onto a curve lying between the nullclines and passing through the fixed point(s). This in turn may be reasonably approximated by the vectorfield linearized at the central, stable *or* unstable, equilibrium and projected onto its weak stable or unstable eigenvector; in the latter case, we implicitly

assume that the preparatory period is short enough that solutions cannot spread too far from the saddle point, and bimodal distributions do not have time to develop (cf. Figs. 3(a,b) and 4(a)).

Since any sum of independent processes of the form $aW_{1,t} + \sqrt{1-a^2}W_{2,t}$ is again Wiener, random forcing along the relevant eigenvector may be represented by the single Wiener process W_t. We take the initial distribution to be the delta function $\delta(0,0)$; the results can be generalized to Gaussian initial data of given mean and variance. Our reduced problem therefore becomes the Ornstein-Uhlenbeck (OU) process

$$du^{\mathrm{P}} = \left(\lambda_{\mathrm{P}} u^{\mathrm{P}}\right) dt + \sigma dW_t \ , \tag{19}$$

with corresponding forward Komolgorov equation

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial u^{\mathrm{P}}}(\lambda_{\mathrm{P}} u^{\mathrm{P}} p) + \frac{\sigma^2}{2}\frac{\partial^2 p}{\left(\partial u^{\mathrm{P}}\right)^2} \ , \ p(u^{\mathrm{P}}; t=0) = \delta({u_0}^{\mathrm{P}}) \ , \tag{20}$$

where λ_{P} is the weak stable or unstable eigenvalue for the preparatory cycle (P denotes preparatory), u defines the distance along the corresponding eigenvector $\mathbf{u}^{\mathrm{P}}$, with $u = 0$ at the fixed point $({x_1}^{\mathrm{P}}, {x_2}^{\mathrm{P}})$, and we allow ${u_0}^{\mathrm{P}} \neq 0$ to include asymmetric biases $b_1 \neq b_2$, for which the initial condition $(x_1, x_2) = (0,0)$ does not coincide with the fixed point. The OU process also arises (in a different manner) in [28]. Eqn. (20) is solved by

$$p(u;t) = \mathcal{N}\left({u_0}^{\mathrm{P}} e^{\lambda_{\mathrm{P}} t}, \frac{\sigma^2}{2\lambda_{\mathrm{P}}}\left(e^{2\lambda_{\mathrm{P}} t} - 1\right)\right) \ , \tag{21}$$

where

$$\mathcal{N}(\mu, \nu^2) = \frac{1}{\sqrt{2\pi\nu^2}} \exp\left(\frac{-(x-\mu)^2}{2\nu^2}\right) \tag{22}$$

denotes the Gaussian (normal) density with mean μ and variance ν^2. In the monostable case ($\lambda_{\mathrm{P}} < 0$), $\mu \to 0$ and p converges on the equilibrium distribution $\mathcal{N}(0, -\sigma^2/2\lambda_{\mathrm{P}})$; in the bistable case ($\lambda_{\mathrm{P}} > 0$), μ increases exponentially and p 'flattens out.' The resulting density at stimulus presentation $p_{\mathrm{T}}(v)$ (T denotes trial) may be estimated from $p(u, \tau_{\mathrm{P}})$ as described below.

With stimuli present ($\rho_j \neq 0$), the equilibrium shifts and one may approximate the dynamics by considering the evolution of $p_{\mathrm{T}}(v + v_0)$ under the drift field established by projection of the trial vectorfield onto the weak stable or unstable eigenvector $\mathbf{v}$ of the unique stable (corresponding to the decision for unambiguous stimuli) or central unstable (for ambiguous stimuli) equilibrium, with eigenvalue λ_{T}. Here the additional shift v_0 accounts

for the fact that the origin of the new ($\mathbf{v}$) coordinate coincides with the trial fixed point $({x_1}^{\mathrm{T}}, {x_2}^{\mathrm{T}})$, which differs from the preparatory phase equilibrium. Thus we must solve

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial v}(\lambda_{\mathrm{T}} v p) + \frac{\sigma^2}{2}\frac{\partial^2 p}{\partial u^2} \ , \ p(v; t=0) = p_{\mathrm{T}}(v + v_0) \ . \tag{23}$$

Letting $v_{\theta j}$ denote the points where the eigenvector intersects the thresholds x_θ, we compute the probabilities $P_j(t)$ that the j^{th} threshold has been passed at time t as

$$P_1^f(t) = \int_{v_{\theta 1}}^{\infty} p(v;t)\, dv \ , \ P_2^f(t) = \int_{-\infty}^{v_{\theta 2}} p(v;t)\, dv \ , \tag{24}$$

where we have used the convention that v increases as x_1 increases. The associated fluxes of probability across decision thresholds, $R_i^f(t)$ (the superscript standing for flux), is found from

$$R_i^f(t) = \frac{d}{dt}(P_i^f(t)) \ . \tag{25}$$

In making this one-dimensional reduction, we implicitly assume that the slow manifolds (relevant eigenvectors) contain phase space regions below the threshold $x_2 = x_\theta$ (resp. to the left of $x_1 = x_\theta$).

These functions may be computed as follows, for both the bistable (ambiguous stimulus, $\lambda_{\mathrm{T}} > 0$) and the monostable (unambiguous stimulus, $\lambda_{\mathrm{T}} < 0$) cases. The initial condition for Eqn. (23) derives from the solution (21) at the close of the preparatory cycle; if we *assume* that the relevant preparatory and trial eigenvectors $\mathbf{u}^{\mathrm{P}}$ and $\mathbf{v}$ may be approximated as parallel, we have

$$p_{\mathrm{T}}(v + v_0) = \mathcal{N}\left(u_0 e^{\lambda_{\mathrm{P}}\tau_{\mathrm{P}}} + v_0, \frac{\sigma^2}{2\lambda_{\mathrm{P}}}\left(e^{2\lambda_{\mathrm{P}}\tau_{\mathrm{P}}} - 1\right)\right) \equiv \mathcal{N}(\mu_0, \nu_0^2) \ , \tag{26}$$

and Eqn. (23) is solved by

$$p(v,t) = \mathcal{N}\left(\mu_0 e^{\lambda_{\mathrm{T}} t},\ \nu_0^2 e^{2\lambda_{\mathrm{T}} t} + \frac{\sigma^2}{2\lambda_{\mathrm{T}}}\left(e^{2\lambda_{\mathrm{T}} t} - 1\right)\right) \equiv \mathcal{N}\left(\mu(t), \nu^2(t)\right) . \tag{27}$$

Using (27) in (24) and performing the changes of variables

$$\xi = \frac{v - \mu(t)}{\sqrt{2\nu^2(t)}} \ \text{ and } \ \Delta_j(t) = \frac{v_{\theta j} - \mu(t)}{\sqrt{2\nu^2(t)}}, \tag{28}$$

we find

$$\begin{aligned} P_1^f(t) &= \frac{1}{\sqrt{\pi}} \int_{\Delta_1(t)}^{\infty} e^{-\xi^2}\, d\xi \\ &= \frac{1}{2}\left[1 - \text{sign}\left(\Delta_1(t)\right) \text{erf}(|\Delta_1(\text{t})|)\right]. \end{aligned} \tag{29}$$

Similarly, cf. [28],

$$\begin{aligned} P_2^f(t) &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\Delta_2(t)} e^{-\xi^2}\, d\xi \\ &= \frac{1}{2}\left[1 + \text{sign}\left(\Delta_2(t)\right) \text{erf}(|\Delta_2(\text{t})|)\right]. \end{aligned} \tag{30}$$

Putting these results into Eqn. (25), we obtain

$$R_1^f(t) = -\frac{1}{\sqrt{\pi}}\left(-e^{-\Delta_1(t)^2}\Delta_1'(t)\right) \;\; ; \;\; R_2^f(t) = \frac{1}{\sqrt{\pi}}\left(e^{-\Delta_2(t)^2}\Delta_2'(t)\right) \;\; , \tag{31}$$

which reduce to

$$\begin{aligned} R_1^f(t) &= -\frac{\left(\sigma^2\mu(t) - v_{\theta 1}\text{e}^{2\lambda_T t}(2\lambda_T\nu_0^2 + \sigma^2)\right)}{\sqrt{\pi}(2\nu^2(t))^{3/2}} \exp\left(-\frac{[v_{\theta 1} - \mu(t)]^2}{2\nu^2(t)}\right) \\ R_2^f(t) &= \frac{\left(\sigma^2\mu(t) - v_{\theta 2}\text{e}^{2\lambda_T t}(2\lambda_T\nu_0^2 + \sigma^2)\right)}{\sqrt{\pi}(2\nu^2(t))^{3/2}} \exp\left(-\frac{[v_{\theta 2} - \mu(t)]^2}{2\nu^2(t)}\right) \frac{\mathbf{v}}{\|\mathbf{v}\|} \end{aligned} \tag{32}$$

with $\mu(t)$ and $\nu(t)$ given as in Eqn. (27), and mean and variance:

$$\langle R^f(t)\rangle = \int_0^{\infty} tR^f(t)\, dt \;\; \text{and} \;\; \text{var}(R^f(t)) = \int_0^{\infty} t^2 R^f(t)\, dt\,. \tag{33}$$

Fig. 5 shows an example of these distributions. Here, we replace the unambiguous stimulus specified in the standard parameter set with the ambiguous values $.55 = \rho_2 > \rho_1 = .45$ to produce nonzero curves $P_j^f(t)$, $R_j^f(t)$ for j both 1 and 2 and also set $b_j = i_0 = 0$. Note that the analytical approximations generally capture the correct form of $P_j^f(t)$ and $R_j^f(t)$, but are significantly shifted from the results of the 2-D simulations; this is largely due to our assumption that solutions collapse in negligible time onto $\mathbf{u}^{\text{P}}$. Corrections developed in the next section address this and other shortcomings.

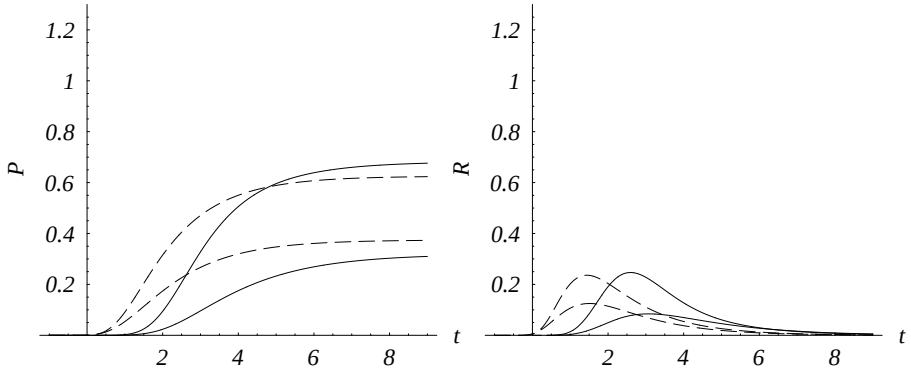


Figure 5: Numerical simulations (solid lines) and analytical approximations (dashed lines) for ambiguous stimulus $.55 = \rho_2 > \rho_1 = .45$. (left) probabilities $P_j^f(t)$ and (right) crossing fluxes $R_j^f(t)$. Upper curves correspond to $j = 2$ (ambiguously correct decision), lower curves to $j = 1$.

3.4.2 Modifications to general method for salient stimuli

In this section, we develop modifications appropriate for the case of strongly salient stimuli (e.g. $|\rho_2 - \rho_1| \geq\sim .25$), a situation treated in the remainder of the paper. We restate three of the assumptions used in developing Eqns. (32): (*i*) the linearization of Eqn. (1) about the relevant fixed point adequately approximates the vector field in the regions affecting the decision dynamics, (*ii*) computing the evolution of $p(x_1, x_2, t; p_0)$ only along the eigenvectors $\mathbf{u}^{\mathrm{P}}$ and $\mathbf{v}$ is sufficient to characterize the two-dimensional distribution and (*iii*) orthogonal projection of the distribution $p(u, \tau_{\mathrm{P}})$ from $\mathbf{u}^{\mathrm{P}}$ onto the (not necessarily parallel) eigenvector $\mathbf{v}$ introduces only a small error into $R_i^f(t)$.

To simplify the resulting expressions, we chose to accept (*iii*) under the assumption that the salient trial dynamics are dominated by the vectorfield linearized at the trial fixed point, which is generally a stable 'star' node (cf., Eqn. (36) below). This implies that $p(x_1, x_2, t)$ should contract onto $\mathbf{v}$ as its mean progresses along this eigenvector toward the decision threshold, so that if the contraction is sufficiently strong the mass of $p(x_1, x_2, \tau_{\mathrm{P}})$ in "bins" perpendicular to $\mathbf{v}$ would cross the threshold roughly simultaneously. Fig. 6 shows a comparison of the (numerically computed) values of $p(x_1, x_2, \tau_{\mathrm{P}})$ along $\mathbf{v}$ with the orthogonal projection of this distribution onto $\mathbf{v}$ at equally spaced points. The similarity between the curves in Fig. 6 supports our

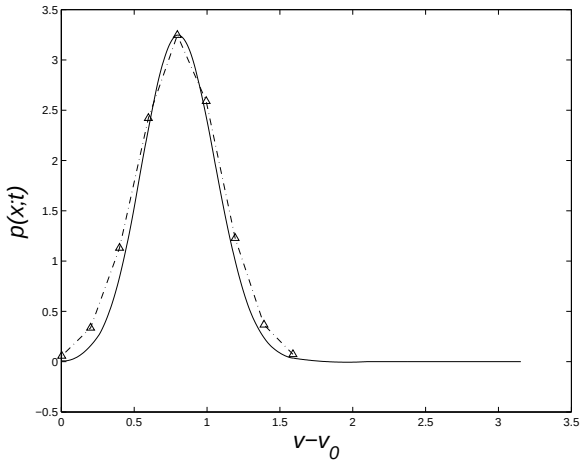


Figure 6: Similarity of the values of $p(x_1, x_2, \tau_{\mathrm{P}})$ along $\mathbf{v}$ (solid line) with orthogonal projections of the distribution onto $\mathbf{v}$ (dashed line). To facilitate comparison, the curves were scaled to have the same maximum.

acceptance of assumption (*iii*) as a reasonable approximation.

In the presence of a salient stimulus, assumptions (*i*) and (*ii*) do not generally hold. In particular, (*i*) may be violated when the relevant evolution during the trial phase occurs far away from the fixed point (i.e. if the fixed point is far outside of the decision thresholds). This is generally the case for strongly salient stimuli, but not for ambiguous stimuli, when the central unstable fixed point is generally near the main diagonal of the phase plane (cf. Figs. 2 and 4). Moreover, (*ii*) may be false when τ_{P} is small, as further explained below.

The first modification to the general method involves the choice of the eigenvector $\mathbf{v}$ in Eqn. (23), and accounts for fact that the mean of $p(x_1, x_2, t)$ approaches but does not completely reach $\mathbf{u}^{\mathrm{P}}$ during the finite period τ_{P} of the preparatory phase. Fig. 7 illustrates this for various values of τ_{P} up to $\tau_{\mathrm{P}} = 1$ (which represents a preparatory period of approximately the same duration as a typical RT in model time units). If τ_{P} is sufficiently small, the resulting effects on the RT distributions calculated in Eqn. (31) and other measures may be significant.

This can be partially corrected for by calculating the linearized *two*-dimensional prediction $(\tilde{x_1}^{\mathrm{P}}, \tilde{x_2}^{\mathrm{P}})$ for the position of the mean at τ_{P}. First, we define the following terms: when the central unstable or single stable fixed point for the preparatory phase is $({x_1}^{\mathrm{P}}, {x_2}^{\mathrm{P}})$, the initial condition $(x_1(0), x_2(0)) = (0, 0)$ at $t = 0$ corresponds to u_0^{Q} and u_0^{P} along $\mathbf{u}^{\mathrm{P}}$ and the stable eigenvector $\mathbf{u}^{\mathrm{Q}}$ for the preparatory phase respectively. These values are given by:

$$u_0^{\mathrm{Q}} \;=\; \sqrt{1 + (m^{\mathrm{Q}})^2} \left(\frac{m^{\mathrm{P}} {x_1}^{\mathrm{P}} - {x_2}^{\mathrm{P}}}{m^{\mathrm{Q}} - m^{\mathrm{P}}} \right)$$

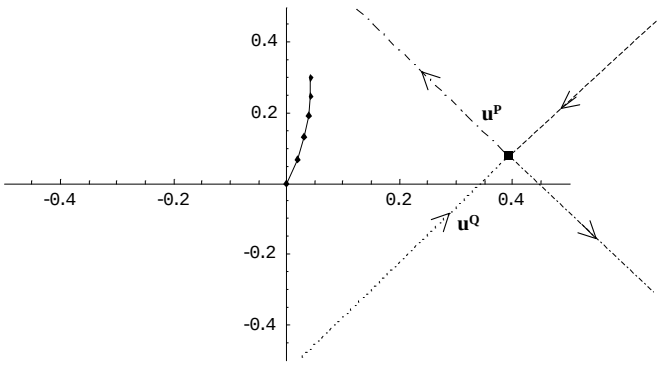


Figure 7: A plot of the mean of $p(x_1, x_2; t = \tau_{\mathrm{P}})$ for various values of τ_{P}. The asymptotic approach of the mean approach to the unstable preparatory cycle eigenvector is addressed by the "first" improvement to the analytical model detailed in the text. Consecutive dots are separated by timesteps $\delta t = .2$, out to a maximum of $\tau_{\mathrm{P}} = 1$, and the standard parameter set with stimulus history AAAA is used.

$$u_0^{\mathrm{P}} \;=\; \sqrt{1 + (m^{\mathrm{P}})^2} \left(\frac{{x_2}^{\mathrm{P}} - m^{\mathrm{Q}} {x_1}^{\mathrm{P}}}{m^{\mathrm{Q}} - m^{\mathrm{P}}} \right) \quad , \tag{34}$$

where m^{P} and m^{Q} are the slopes of $\mathbf{u}^{\mathrm{Q}}$ and $\mathbf{u}^{\mathrm{P}}$ in the (x_1, x_2) phase plane. At the end of the preparatory cycle, the mean of the distribution is then located at the point

$$\begin{aligned} \tilde{{x_1}^{\mathrm{P}}} &= \frac{\left({x_2}^{\mathrm{P}} - m^{\mathrm{Q}} {x_1}^{\mathrm{P}} \right) \mathrm{e}^{\tau_{\mathrm{P}} \lambda_{\mathrm{P}}} + \left(m^{\mathrm{P}} {x_1}^{\mathrm{P}} - {x_2}^{\mathrm{P}} \right) \mathrm{e}^{\tau_{\mathrm{P}} \lambda_{\mathrm{Q}}}}{m^{\mathrm{Q}} - m^{\mathrm{P}}} + {x_1}^{\mathrm{P}} \\ \tilde{{x_2}^{\mathrm{P}}} &= \frac{m^{\mathrm{P}} \left({x_2}^{\mathrm{P}} - m^{\mathrm{Q}} {x_1}^{\mathrm{P}} \right) \mathrm{e}^{\tau_{\mathrm{P}} \lambda_{\mathrm{P}}} + m^{\mathrm{Q}} \left(m^{\mathrm{P}} {x_1}^{\mathrm{P}} - {x_2}^{\mathrm{P}} \right) \mathrm{e}^{\tau_{\mathrm{P}} \lambda_{\mathrm{Q}}}}{m^{\mathrm{Q}} - m^{\mathrm{P}}} + {x_2}^{\mathrm{P}} \end{aligned} \tag{35}$$

In the case of unambiguous stimuli, the Jacobian of Eqns. (1) for the trial phase evaluated at the unique stable fixed point $({x_1}^{\mathrm{T}}, {x_2}^{\mathrm{T}})$ is approximately given by:

$$J_{\mathrm{T}} \simeq \begin{bmatrix} -k & 0 \\ 0 & -k \end{bmatrix} = -kI, \tag{36}$$

so under this approximation any nonzero vector in $\Re^2$ is a stable eigenvector with eigenvalue $\lambda_{\mathrm{T}} = -k$ for the trial phase. Hence, $\mathbf{v}$ may be chosen as the eigenvector connecting $({x_1}^{\mathrm{T}}, {x_2}^{\mathrm{T}})$ with $(\tilde{{x_1}^{\mathrm{P}}}, \tilde{{x_2}^{\mathrm{P}}})$; this generally results in an improvement in accuracy over the eigenvector connecting $({x_1}^{\mathrm{T}}, {x_2}^{\mathrm{T}})$ with $({x_1}^{\mathrm{P}}, {x_2}^{\mathrm{P}})$. In this case, the initial condition (26) for Eqn. (23) is replaced

by

$$p(v;t=0) = \mathcal{N}\left(\tilde{v}_0, \frac{\sigma^2}{2\lambda_{\mathrm{P}}}\left(\mathrm{e}^{2\lambda_{\mathrm{P}}t}-1\right)\right) \quad , \tag{37}$$

where

$$\tilde{v}_0 = \left[({x_1}^{\mathrm{T}} - \tilde{x_1}^{\mathrm{P}})^2 + ({x_2}^{\mathrm{T}} - \tilde{x_2}^{\mathrm{P}})^2\right]^{\frac{1}{2}} \tag{38}$$

and we retain the variance predicted in Eqn. (26). Denoting by $(x_\theta, {x_2}^\theta)$ and (x_1^θ, x_θ) the intersections of the stable eigenvector with the thresholds $x_j = x_\theta$ ($j = 1, 2$ respectively), we obtain the threshold values along $\mathbf{v}$

$$v_{\theta 1} = \left[({x_1}^{\mathrm{T}} - x_\theta)^2 + ({x_2}^{\mathrm{T}} - {x_2}^\theta)^2\right]^{\frac{1}{2}} \tag{39}$$

$$v_{\theta 2} = \left[({x_1}^{\mathrm{T}} - {x_1}^\theta)^2 + ({x_2}^{\mathrm{T}} - x_\theta)^2\right]^{\frac{1}{2}} \quad . \tag{40}$$

A second correction addresses the approximation of replacing the projection of the vector field (X_1, X_2) onto $\mathbf{v}$ by its linearization at $({x_1}^{\mathrm{T}}, {x_2}^{\mathrm{T}})$. Fig. 8 shows a typical projection in the case of salient stimuli, and demonstrates that the vector field may indeed be approximated as linear in the 'post-threshold' region 1. However, in the pre-threshold region 2, a constant vectorfield approximation appears to be more appropriate, with the constant value V determined by averaging over the relevant region along v or simply by the value of the projection at an appropriate point along v. Techniques do exist (e.g. [10]) to compute the approximate evolution for piecewise linear vectorfields such as that spanning regions 1 and 2; however, to simplify the analysis here we will consider using one of *either* the constant or linearized velocity profiles.

Since the initial distribution $p_{\mathrm{T}}(v + v_0)$ for the trial phase is largely supported between the thresholds (Fig. 8) and we are primarily concerned with hitting time distributions $R_j^h(t)$ (independent of post-threshold dynamics), a constant vectorfield approximation seems most appropriate for the free response protocol. For the forced response protocol or the trial-dependent updating of model parameters (e.g. Eqn.(3)), in which post-threshold dynamics may be significant, the linear (1) and constant (2) regions may be effectively 'averaged' by modifying the linear Ornstein-Uhlenbeck formulation via $\lambda_{\mathrm{T}} \to \psi\lambda_{\mathrm{T}}$, $\psi > 1$. Here, where w.l.o.g. stimulus "2" is assumed correct, we choose

$$\psi = \frac{1}{\lambda_{\mathrm{T}} v_{\theta 2}} \left(X_1(\mathbf{x}^\psi), X_2(\mathbf{x}^\psi)\right) \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|} \quad , \tag{41}$$

where $\mathbf{x}^\psi$ is the selected point along $\mathbf{v}$.

To accommodate a constant vector field approximation with velocity V , the results derived above for the Ornstein-Uhlenbeck case may be modified as follows. Eqn. (19) becomes the equation for constant drift (i.e., uniform velocity) Brownian motion (CDBM),

$$dv = V dt + \sigma \, dW_t \ , \tag{42}$$

and Eqn. (23) becomes

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial v}(Vp) + \frac{\sigma^2}{2}\frac{\partial^2 p}{\partial u^2} \ . \tag{43}$$

The latter equation is solved by

$$p(v;t) = \mathcal{N}\left(\mu(t), \nu^2(t)\right) \quad , \tag{44}$$

where

$$\mu(t) = Vt + \tilde{v}_0 \quad \nu^2(t) = \nu_0^2 + \sigma^2 t \ . \tag{45}$$

From these equations, $P_j(t)$ may be computed via Eqns. (29)-(30) and $R_j^f(t)$ via Eqn. (31); this yields, in place of Eqns. (32) :

$$\begin{aligned} R_1^f(t) &= -\frac{\sigma^2\left(\tilde{v}_0 - v_{\theta 1} - Vt\right) - 2\nu_0^2 V}{2\sqrt{2\pi}\left(\nu_0^2 + \sigma^2 t\right)^{3/2}} \exp\left[-\frac{\left(\tilde{v}_0 - v_{\theta 1} + Vt\right)^2}{2\left(\nu_0^2 + \sigma^2 t\right)}\right] \quad , \\ R_2^f(t) &= \frac{\sigma^2\left(\tilde{v}_0 - v_{\theta 2} - Vt\right) - 2\nu_0^2 V}{2\sqrt{2\pi}\left(\nu_0^2 + \sigma^2 t\right)^{3/2}} \exp\left[-\frac{\left(\tilde{v}_0 - v_{\theta 2} + Vt\right)^2}{2\left(\nu_0^2 + \sigma^2 t\right)}\right] \quad . \end{aligned} \tag{46}$$

We discuss estimates of V and ψ in Section 3.5.

As quantified below, the CDBM model is generally a superior approximation in the case of salient stimuli; other aspects of the relationship between CDBM and OU models are discussed in [28].

3.4.3 Relationship between fluxes in the forced-response protocol and reaction times in the free-response protocol

Under appropriate assumptions, the fluxes $R_j^f(t)$ found in the previous section may be used to represent reaction time densities in the free-response protocol. Specifically, we assume that (i) the stimuli are sufficiently unambiguous so that $P_1^f(t)$ and $R_1^f(t)$ are negligible (assuming again w.l.o.g. that stimulus 2 is correct) and (ii) the projection of the drift vectorfield normal to the correct decision threshold is sufficiently positive, where "sufficiently" will be made precise in what follows.

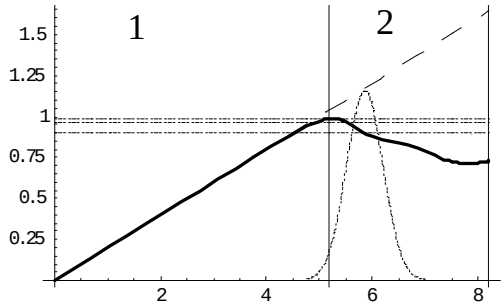


Figure 8: Projection of the vectorfield (X_1, X_2) onto $\mathbf{v}$ for standard parameter set with AAAA stimulus history, with the sign switched for plotting and the regions '1' and '2' discussed in the text indicated. The lefthand of the two vertical gridlines represents the location of the correct decision threshold, the righthand gridline that of the incorrect threshold. The superimposed Gaussian shows the position of $p(v;t)$ at the end of the preparatory phase. Three horizontal lines show, in decreasing order, the values of $\mathbf{X}(\mathbf{x_V}) \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|}$ for $\mathbf{x_V} = (x_{\theta 1}, x_\theta)$, $\mathbf{x_V} = (\tilde{x_1^{\rm P}} + x_{\theta 1}, \tilde{x_2^{\rm P}} + x_\theta)/2$, and $\mathbf{x_V} = (\tilde{x_1^{\rm P}}, \tilde{x_2^{\rm P}})$ (cf., Section 3.5).

The key difference between the forced- and free-response protocols is that, in the latter, individual realizations of Eqn. (1) must be removed from calculations of reaction times at first crossing of either threshold. The corresponding density of $v_{\theta j}$-hitting times for Brownian motion with constant drift V and initial condition $\tilde{v}_0$ may be computed as

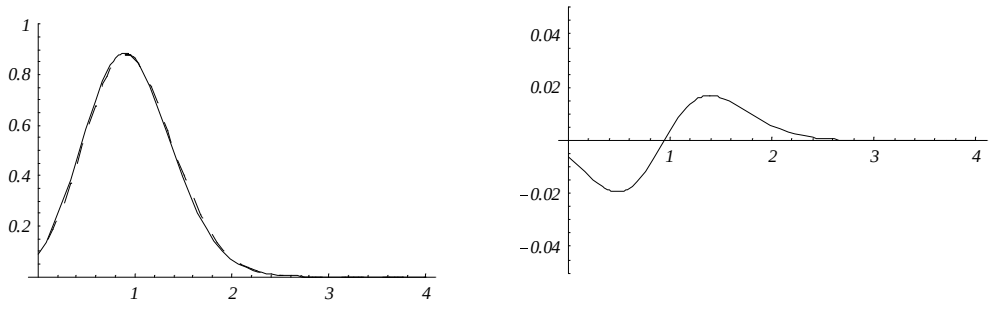
$$R^h(t) = \frac{\sigma^2\left(\tilde{v}_0 - v_{\theta j}\right) - \nu_0^2 V}{\sqrt{2\pi}\left(\nu_0^2 + \sigma^2 t\right)^{3/2}} \exp\left[-\frac{\left(\tilde{v}_0 - v_{\theta j} + Vt\right)^2}{2\left(\nu_0^2 + \sigma^2 t\right)}\right] \quad , \tag{47}$$

using the optional sampling theorem (cf., [18]). Comparing this expression with the time-dependent probability flux calculated in Eqn. (46), we find explicitly the difference between barrier hitting time distributions and fluxes:

$$R^f(t) - R^h(t) = -\frac{\sigma^2\left(\tilde{v}_0 - v_{\theta j} + Vt\right) \exp\left(-\frac{(\tilde{v}_0 - v_{\theta j} + Vt)^2}{2(\nu_0^2 + \sigma^2 t)}\right)}{2\sqrt{2\pi}\left(\nu_0^2 + \sigma^2 t\right)^{3/2}} \tag{48}$$

which for fixed time scales with $\exp(-V^2 t)$.

A comparison of flux probabilities $R^f(t)$ and hitting time distributions $R^h(t)$ is given in Fig. 9 for values of ν_0^2, $v_{\theta 2}$, and $\tilde{v_0}$ from the equal bias case



a) b)

Figure 9: Comparison of flux probabilities $R_f(t)$ (dashed line) and hitting time distributions $R_h(t)$ (solid line) for 1-D, constant-drift Brownian motion. Parameters are drawn from the equal bias case of the standard parameter set, and a typical value of $V \approx -0.9$ was used. a) Overlay of the two distributions; b) difference $R_f(t) - R_h(t)$.

of the standard parameter set. The L^1 norm of $R^f - R^h$ was found to be approximately .026, or 2.6% of the norm of the probability density. For the remainder of the paper we will assume that this norm remains small. Hence, while we continue to compute probability fluxes, we will frequently drop the superscripts f and h and refer to our results as reaction times of the free-response protocol, tacitly assuming that $R(t) = R^f(t) \approx R^h(t)$. Integrating this relationship, we have

$$\mathrm{P}\left\{\mathrm{T}_{x_\theta} < t\right\} = \mathrm{P}\left\{x_j(t) > x_\theta\right\} + \mathcal{O}\left(\exp(-V^2 t)\right) \ , \tag{49}$$

where for a particular realization of the processes $W_{j,t}$, $\mathrm{P}\left\{x_j(t) > x_\theta\right\}$ is the probability that the activity of the j^{th} unit has exceeded threshold at time t and $\mathrm{P}\left\{\mathrm{T}_{x_\theta} < t\right\}$ is the probability that the hitting time of x_j has already occurred.

To 'close' the analytical expressions (32, 46), it remains to compute or approximate the values of $\tilde{v_0}$, $v_{\theta j}$, V, and ν_0^2, to which we now turn.

3.4.4 Approximate closed-form expressions for salient stimuli

Depending on the form of the activation function $f(x)$, the solution to the fixed point equations (14), and hence the values $v_{\theta j}$, v_0, $\lambda_{\mathrm{P,Q,T}}$, and u_0, may require numerical techniques. We therefore make the piecewise-linear

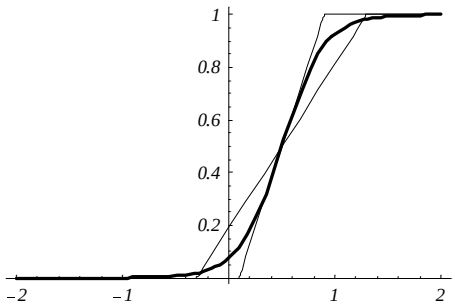


Figure 10: Comparison of logistic and piecewise linear activation functions for the standard parameter set. The linear function with the smaller slope ($1/2 \max[f']$) is used to compute eigenvalues: see text following Eqn. (57).

approximation

$$f(x;g,b) \approx \hat{f}(x;g,b) = \begin{cases} 0 & \text{if } x < b - \frac{2}{g} \\ \frac{g}{4}\left(x + \frac{2}{g} - b\right) & \text{if } b - \frac{2}{g} \le x \le b + \frac{2}{g} \\ 1 & \text{if } x > b + \frac{2}{g} \end{cases} \quad . \tag{50}$$

Note that $\hat{f}(x;g,b)$ has either zero or the maximum slope g/4 of $f(x;g,b)$. Approximating thresholds by $x_\theta = \hat{f}^{-1}(\theta)$, we obtain $x_\theta = b + (4\theta - 2)/g$; since the slope of $\hat{f}(x;g,b)$ is less than that of $f(x;g,b)$ for $x < b$ and greater than that of $f(x;g,b)$, for $x > b$, the piecewise linear function typically overestimates x_θ for $\theta < 1/2$ and underestimates x_θ for $\theta > 1/2$. See Fig. 10. In the symmetric case ($b_1 = b_2$), the condition for bistability remains $\beta g > 4k$ and the pitchfork bifurcations occur at

$$\hat{\gamma}_{\rm pf} = b \pm 2/g \quad . \tag{51}$$

We will use Eqn. (50) to obtain closed-form approximations for the quantities calculated in the previous section. Depending on parameter values, the nullclines may intersect at different sections of the piecewise linear activation function, so there are several separate cases to consider. For a strongly unambiguous stimulus $\rho_2 = 0.85$ and the standard parameter set, the nullclines intersect in the region $b - 2/g \le x_j \le b + 2/g$, $j = 1, 2$ during the preparatory phase and $x_2 > b + 2/g$, $x_1 < b + 2/g$ during the trial phase. While we only work through this case in detail, the expressions for similar preparatory cycle parameter values but unambiguous stimuli $\rho_1 > \rho_2$ are nearly identical in form to those calculated below. In other parameter ranges the nullcline intersection patterns differ, but the same methods apply.

Introducing the notation $\gamma_j^{\mathrm{P}} = i_0 + b_j, \gamma_j^{\mathrm{T}} = i_0 + b_j + \rho_j \;\; j = 1, 2$, the location of the central saddle point of the preparatory phase dynamics is given by solving Eqn. (14) with f replaced by $\hat{f}$:

$$
\begin{aligned}
{x_1}^{\mathrm{P}} &= -\frac{2(\gamma_1^{\mathrm{P}} - \gamma_2^{\mathrm{P}})}{\beta g - 4k} + \frac{2(\gamma_1^{\mathrm{P}} + \gamma_2^{\mathrm{P}}) + \beta(bg - 2)}{\beta g + 4k} && (52)\\
{x_2}^{\mathrm{P}} &= \frac{2(\gamma_1^{\mathrm{P}} - \gamma_2^{\mathrm{P}})}{\beta g - 4k} + \frac{2(\gamma_1^{\mathrm{P}} + \gamma_2^{\mathrm{P}}) + \beta(bg - 2)}{\beta g + 4k} \; . && (53)
\end{aligned}
$$

Under the assumption of salient stimuli, the location of the single stable fixed point during the trial phase is given by

$$
{x_1}^{\mathrm{T}} = \frac{\gamma_1^{\mathrm{T}} - \beta}{k} \; , \; {x_2}^{\mathrm{T}} = \frac{\gamma_2^{\mathrm{T}}}{k} \; . \tag{54}
$$

The Jacobian of Eqns. (1) with activation function $\hat{f}(x)$ for the preparatory cycle evaluated at $({x_1}^{\mathrm{T}}, {x_2}^{\mathrm{T}})$ is

$$
J_{\mathrm{P}} = \begin{bmatrix} -k & -\frac{\beta g}{4} \\ -\frac{\beta g}{4} & -k \end{bmatrix}, \tag{55}
$$

which yields $\lambda_{\mathrm{P}} = -k + \beta g/4, \; \lambda_{\mathrm{Q}} = -k - \beta g/4$ with corresponding unstable and stable eigenvectors $(1, -1)$ and $(1, 1)$. It follows from the fact that the stable eigenvector of J_{P} has slope one that

$$
u_0^{\mathrm{P}} = \frac{1}{\sqrt{2}} \left({x_2}^{\mathrm{P}} - {x_1}^{\mathrm{P}} \right) = \frac{2\sqrt{2}(\gamma_1^{\mathrm{P}} - \gamma_2^{\mathrm{P}})}{\beta g - 4k} \tag{56}
$$

and from Eqn. (35) that

$$
\begin{aligned}
{\tilde{x_1}}^{\mathrm{P}} &= \frac{1}{2} \left[\left({x_2}^{\mathrm{P}} - {x_1}^{\mathrm{P}} \right) \mathrm{e}^{\tau_{\mathrm{P}} \lambda_{\mathrm{P}}} - \left({x_1}^{\mathrm{P}} + {x_2}^{\mathrm{P}} \right) \mathrm{e}^{\tau_{\mathrm{P}} \lambda_{\mathrm{Q}}} \right] + {x_1}^{\mathrm{P}} \\
{\tilde{x_2}}^{\mathrm{P}} &= \frac{1}{2} \left[- \left({x_2}^{\mathrm{P}} - {x_1}^{\mathrm{P}} \right) \mathrm{e}^{\tau_{\mathrm{P}} \lambda_{\mathrm{P}}} - \left({x_1}^{\mathrm{P}} + {x_2}^{\mathrm{P}} \right) \mathrm{e}^{\tau_{\mathrm{P}} \lambda_{\mathrm{Q}}} \right] + {x_2}^{\mathrm{P}} \; . && (57)
\end{aligned}
$$

An additional observation is appropriate at this point. Since Eqn. (55) represents the vectorfield linearized at the maximum slope of the 'true' activation function f, which is appropriate only for equal biases, λ_{P} will be overestimated for central fixed points that do not occur on the diagonal. This situation arises for unequal biases b_1 and b_2; an approximate slope if $|b_1 - b_2|$ is sufficiently large is the average between maximum and minimum (zero) slopes of the logistic function, or $\lambda_{\mathrm{P,Q}} \to (1/2)(-k + \beta g/4)$. To simplify calculations, this slope averaging was used to calculate eigenvalues for all parameter sets tested in this paper.

The Jacobian of Eqns. (1) for the trial phase ($\rho_j \neq 0$), evaluated at the unique stable fixed point $({x_1}^{\mathrm{T}}, {x_2}^{\mathrm{T}}) = \left((\gamma_1^{\mathrm{T}} - \beta)/k, \gamma_2^{\mathrm{T}}/k\right)$ is:

$$J_{\mathrm{T}} = \begin{bmatrix} -k & 0 \\ 0 & -k \end{bmatrix} = -kI, \tag{58}$$

so any nonzero vector in $\Re^2$ is a stable eigenvector with eigenvalue $\lambda_{\mathrm{T}} = -k$ during the trial phase. We then project the 1-dimensional distribution along $\mathbf{u}^{\mathrm{Q}}$ without distortion onto the eigenvector $\mathbf{v}$ in the direction $({\tilde{x_1}}^{\mathrm{P}} - {x_1}^{\mathrm{T}}, {\tilde{x_2}}^{\mathrm{P}} - {x_2}^{\mathrm{T}})$, to obtain

$$\tilde{v}_0 = \left(({\tilde{x_1}}^{\mathrm{P}} - {x_1}^{\mathrm{T}})^2 + ({\tilde{x_2}}^{\mathrm{P}} - {x_2}^{\mathrm{T}})^2\right)^{\frac{1}{2}} \ . \tag{59}$$

Using the expressions for ${\tilde{x}_j}^{\mathrm{P}}$ and ${x_j}^{\mathrm{T}}$ derived above, Eqns. (39, 40, 59) may be used to show that

$$\tilde{v}_0 = \frac{\sqrt{\mathrm{A}^2 + \mathrm{B}^2}}{k\left(\beta^2 g^2 - 16k^2\right)} \tag{60}$$

$$v_{\theta 1} = -\frac{\gamma_1^{\mathrm{T}} - \beta - k x_\theta}{k}\sqrt{1 + \left(\frac{B}{A}\right)^2} \tag{61}$$

$$v_{\theta 2} = \frac{\gamma_2^{\mathrm{T}} - k x_\theta}{k}\sqrt{1 + \left(\frac{A}{B}\right)^2} \ , \tag{62}$$

where

$$\begin{aligned} A &= 2\beta g k\left[-D\gamma_1^{\mathrm{P}} + (C-2)\gamma_2^{\mathrm{P}} - \frac{E}{2}\beta(bg-2)\right] \\ &\quad +8k^2\left[-C\gamma_1^{\mathrm{P}} + D\gamma_2^{\mathrm{P}} + \frac{E}{2}\beta(bg-2)\right] + \beta^2 g^2\left(\gamma_1^{\mathrm{T}} - \beta\right) + 16k^2(\beta - \rho_1) \end{aligned} \tag{63}$$

$$\begin{aligned} B &= 2\beta g k\left[(C-2)\gamma_1^{\mathrm{P}} + -D\gamma_2^{\mathrm{P}} - \frac{E}{2}\beta(bg-2)\right] + \\ &\quad 8k^2\left[D\gamma_1^{\mathrm{P}} - C\gamma_2^{\mathrm{P}} + \frac{E}{2}\beta(bg-2)\right] + \beta^2 g^2\left(\gamma_2^{\mathrm{T}} - \beta\right) - 16k^2\rho_2 \end{aligned} \tag{64}$$

$$C = \mathrm{e}^{\tau_{\mathrm{P}}\lambda_{\mathrm{P}}} + \mathrm{e}^{\tau_{\mathrm{P}}\lambda_{\mathrm{Q}}}$$

$$D = \mathrm{e}^{\tau_{\mathrm{P}}\lambda_{\mathrm{P}}} - \mathrm{e}^{\tau_{\mathrm{P}}\lambda_{\mathrm{Q}}}$$

$$E = 1 - \mathrm{e}^{\tau_{\mathrm{P}}\lambda_{\mathrm{Q}}} \ . \tag{65}$$

For the 'AAAA' case introduced above, we compared the values found in Eqns. (52) - (62) using $\hat{f}(x)$ with those found using $f(x)$; the results are summarized in Table 2.

	logistic	piecewise linear
x_1^{P}	0.393	0.372
x_2^{P}	0.0771	0.191
x_1^{T}	-2.20	-2.20
x_2^{T}	5.713	5.713
$v_{\theta 1}$	8.20	7.94
$v_{\theta 2}$	5.17	5.29
u_0^{P}	-.318	-.128
v_0	5.86	6.00
λ^{P}	.361	.369
λ^{Q}	-.761	-.569
λ^{T}	-.2	-.2

Table 2: Comparison of input values for the analytical model calculated using response functions f and $\hat{f}$. Note that the piecewise linear estimate of $\lambda_{\mathrm{P,Q}}$ is obtained by taking the slope of $\hat{f} = 1/2 \max[f']$ as described in the text following Eqn. (57).

As will be further discussed in Section 3.5, the value of V in Eqn. (46) is the projection $\mathbf{X}(\mathbf{x}^{\mathbf{V}}) \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|}$, where $\mathbf{x}^{\mathbf{V}}$ is the location at which the drift vectorfield $\mathbf{X}$ is to be evaluated. This may be written as

$$V = \frac{\dot{x_1}(x_1^V) + \dot{x_2}(x_2^V)\tilde{m_T}}{\sqrt{1 + \tilde{m_T}^2}} \quad , \tag{66}$$

where $\tilde{m_T} = A/B$ is the slope of the relevant trial phase eigenvector. For example, if we take $\mathbf{x}^{\mathbf{V}} = (\tilde{x_1^{\mathrm{P}}}, \tilde{x_2^{\mathrm{P}}})$ and assume equal biases $b_1 = b_2$ (which gives $\tilde{x_1^{\mathrm{P}}} = \tilde{x_2^{\mathrm{P}}} \equiv \tilde{x^{\mathrm{P}}}$), Eqn. (66) becomes

$$V = \frac{\left(-k\tilde{x^{\mathrm{P}}} - \beta f(\tilde{x^{\mathrm{P}}})\right)(1 + B/A) + \gamma_1^{\mathrm{T}} + \gamma_2^{\mathrm{T}} B/A}{\sqrt{1 + (B/A)^2}} \quad . \tag{67}$$

For other parameter sets that may be treated as perturbations from this symmetric case, a useful approximation may be to hold V constant at this value.

We now obtain an explicit approximation for the median of the reaction time distribution in the Ornstein-Uhlenbeck approximation. In the case of highly salient stimuli with alternative 2 correct, the error rate $\max\{P_1(t) : 0 \leq t \leq \tau_{\mathrm{P}} + \tau_{\mathrm{T}}\}$ is in general sufficiently low that we can make the approximation $P_2(t) \to 1$ as $t \to \infty$. Hence, the median of $R(t)$ occurs when

$P_2(t) = 1/2$; Eqn. (30) shows that this occurs where $\Delta_2(t) = 0$. In the Ornstein-Uhlenbeck approximation, this implies

$$\text{median}(R(t)) = \frac{1}{\lambda_{\mathrm{T}}} \ln\left(\frac{v_{\theta 2}}{\mu_0}\right) = \frac{1}{\lambda_{\mathrm{T}}} \ln\left(\frac{v_{\theta_2}}{u_0^{\mathrm{P}} \exp(\lambda_{\mathrm{P}} \tau_{\mathrm{P}}) + v_0}\right). \tag{68}$$

Using Eqns. (62), (56), and (60), this equation becomes

$$\text{median(R(t))} = \frac{1}{k}\left(\ln\left[\frac{\sqrt{2}k(\gamma_1^{\mathrm{P}} - \gamma_2^{\mathrm{P}})\mathrm{e}^{\tau_{\mathrm{P}}\left(-k + \frac{\beta g}{4}\right)}}{\sqrt{1 + (A/B)^2}} + \frac{\beta B}{\beta g + 4k}\right] - \ln[(\gamma_2^{\mathrm{T}} - k x_\theta)(\beta g - 4k)]\right) \tag{69}$$

If the correction term ψ discussed in the previous section is included, this expression is multiplied by $1/\psi$. Finally, in the constant-velocity approximation, $\Delta_2(t) = 0$ yields

$$\begin{aligned} \text{median}(R(t)) &= \frac{v_{\theta 2} - \tilde{v}_0}{V} \\ &= \frac{\sqrt{A^2 + B^2}}{Vk}\left[\frac{\gamma_2^{\mathrm{T}} - k x_\theta}{B} - \frac{1}{\beta^2 g^2 - 16k^2}\right] . \end{aligned} \tag{70}$$

We note that closed form expressions for $R(t)$ could also be derived via Eqn. (32) or Eqn. (46) and used to calculate moments of the RT distribution.

3.5 Accuracy of the analytical approximations

Reaction time distributions obtained using the Ornstein-Uhlenbeck (OU) and constant-drift Brownian motion (CDBM) approximations with logistic response functions (Eqns. (32, 46)) to Eqns. (1) were compared with 2-D numerical simulations for the AAAA, AAAR, and equal bias parameter sets given in Table 1. For the OU and CDBM models, three different approximations for the drift vectorfield were examined, each for linear (OU) or constant (CDBM) vectorfields with projections $\mathbf{X}(\mathbf{x_V}) \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|}$ calculated at a different point along $\mathbf{v}$. Specifically, the approximations were tested with $\mathbf{x_V}$ equal to: (i) $(\tilde{x_1^{\mathrm{P}}}, \tilde{x_2^{\mathrm{P}}})$ (the analytically computed maximum of $p(x_1, x_2; \tau_{\mathrm{P}})$), (ii) $(x_{\theta 1}, x_\theta)$ (the point at which the relevant trial phase eigenvector $\mathbf{v}$ crosses the correct decision threshold), and (iii) $(\tilde{x_1^{\mathrm{P}}} + x_{\theta 1}, \tilde{x_2^{\mathrm{P}}} + x_\theta)/2$ (the intermediate point).

Analytically computed reaction time distributions (denoted here by $R^a(t)$) were compared with numerical data via several metrics, for all possible combinations of drift vector field (i.e. OU or CDBM), specification of $\mathbf{x^V}$, and

	Frac. err., mean	Frac. err. variance	$\|R^a - R^n\|_{L^1}$	$D_{KL}(R^a, R^n)$
CDBM, $\mathbf{x^V} = (\tilde{x_1^{\mathrm{P}}}, \tilde{x_2^{\mathrm{P}}})$	-.036	-.36	.34	.23
CDBM, $\mathbf{x^V} = (\tilde{x_1^{\mathrm{P}}} + x_{\theta 1}, \tilde{x_2^{\mathrm{P}}} + x_\theta)/2$	-.13	-.48	.27	.35
CDBM, $\mathbf{x^V} = (x_{\theta 1}, x_\theta)$	-.17	-.52	.26	.46
OU, $\mathbf{x^V} = (\tilde{x_1^{\mathrm{P}}}, \tilde{x_2^{\mathrm{P}}})$	-.12	-.53	.29	.44
OU, $\mathbf{x^V} = (\tilde{x_1^{\mathrm{P}}} + x_{\theta 1}, \tilde{x_2^{\mathrm{P}}} + x_\theta)/2$	-.21	-.62	.30	.75
OU, $\mathbf{x^V} = (x_{\theta 1}, x_\theta)$	-.25	-.65	.36	1.00
CDBM, PW linear $\hat{f}$ $\mathbf{x^V} = (x_{\theta 1}, x_\theta)$	-.20	-.49	.27	.46

Table 3: Metrics of difference between analytical approximations $R^a(t)$ and numerical simulations $R^n(t)$. Here, the fractional error 'frac. err.' is the difference between statistics computed for $R^a(t)$ and $R^n(t)$, normalized by the values obtained for $R^n(t)$.

stimulus history. The first is the magnitude of differences between means and variances of the distributions $R^a(t)$ and mean and variance of the corresponding numerically computed distribution $R^n(t)$. The L^1 norms and the Kullback-Leibler distances (or cross entropies) [11] between $R^a(t)$ and $R^n(t)$ were also computed, where the latter is specified by

$$D_{KL}(R^a, R^v) = \int_{-\infty}^{\infty} R^n(t) \log\left(\frac{R^n(t)}{R^a(t)}\right) \, dt \quad . \tag{71}$$

The KL norm effectively emphasizes errors in distributions' tails, where absolute values and hence L^1 differences are both small.

Averaged over the three cases of stimulus history, the CDBM model with $\mathbf{x^V} = (\tilde{x_1^{\mathrm{P}}}, \tilde{x_2^{\mathrm{P}}})$ gave the smallest values of all of these metrics, except for the L^1 norm; $\mathbf{x^V} = (\tilde{x_1^{\mathrm{P}}}, \tilde{x_2^{\mathrm{P}}})$ was also found to be (here, uniformly) preferable for the OU model. The best approximation in the L^1 sense was given by CDBM with $\mathbf{x^V} = (x_{\theta 1}, x_\theta)$. These and other values are given in Table 3, and Figs. 11-13 show the relevant comparisons of reaction time distributions.

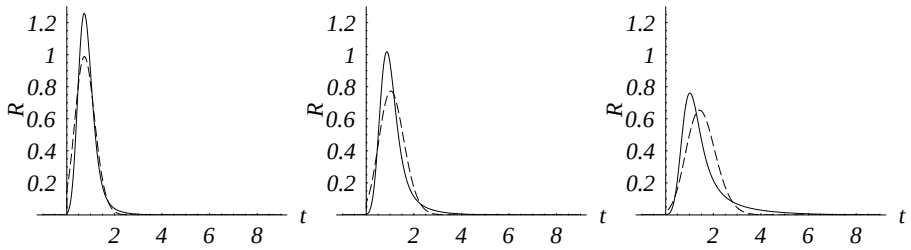


Figure 11: Comparison of $R(t)$ calculated using the 2-D numerical simulations (solid) and the 1-D approximations with logistic response functions for CDBM model (dashed), with $\mathbf{x^V} = (\tilde{x_1^{\rm P}}, \tilde{x_2^{\rm P}})$. For this and Figs. 12 and 13, results are given for the standard parameter set and, from left to right, stimulus history AAAA, equal bias, and stimulus history AAAR.

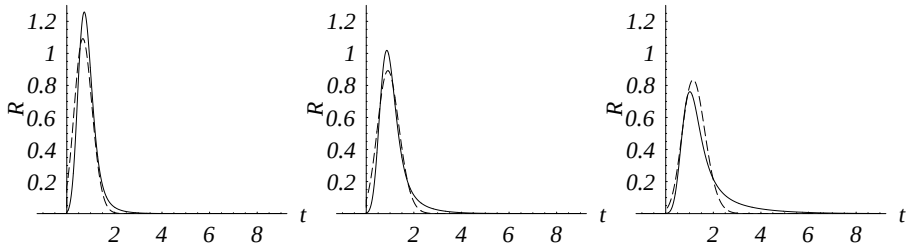


Figure 12: Comparison of $R(t)$ calculated using the 2-D numerical simulations (solid) and the 1-D approximations with logistic response functions for CDBM model (dashed), with $\mathbf{x^V} = (x_{\theta 1}, x_\theta)$.

Fig. 8 reveals why the smallest choice of constant drift velocity, calculated at $\mathbf{x^V} = (\tilde{x_1^{\rm P}}, \tilde{x_2^{\rm P}})$, gives the best values for statistical properties of $R^a(t)$, and also why these approximations generally fail to reproduce the long tails of the numerical simulations. The effective one-dimensional vectorfield is locally *increasing* approaching the threshold from region 2. A linear approximation to such a vectorfield would be equivalent to that obtained from linearization about a fictitious *unstable* fixed point located to the right of the domain of Fig. 8. Eqn. (27) shows that such a vectorfield would result in variance of the Gaussian $p(v;t)$ increasing exponentially in time, resulting in a lesser slope for the trailing edge of the Gaussian and hence a longer tail in $R(t)$. Meanwhile the variance of the constant drift distribution $p(v;t)$ given in Eqn. (45) increases linearly in time, resulting in

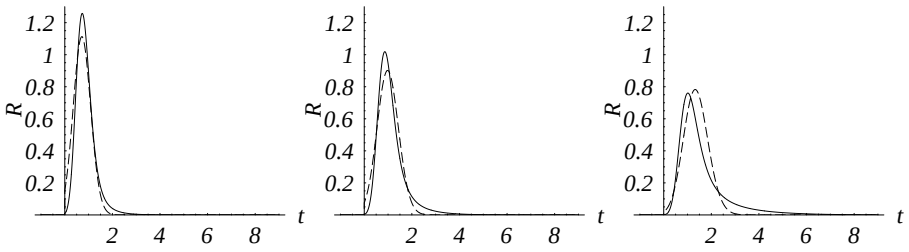


Figure 13: Comparison of $R(t)$ calculated using the 2-D numerical simulations (solid) and the 1-D approximations with logistic response functions for the OU model (dashed), with $\mathbf{x}^{\psi} = (\tilde{x_1^{\mathrm{P}}}, \tilde{x_2^{\mathrm{P}}})$.

shorter tails but retaining some of the structure observed in numerical simulations. By contrast, the Ornstein-Uhlenbeck approximation derived from linearization about the *stable* fixed point $(x_1^{\mathrm{T}}, x_2^{\mathrm{T}})$ results in a *decreasing* variance in time, likely contributing to yet poorer matches for the tails of $R^n(t)$.

In spite of these shortcomings, the analytical approximations with logistic response functions offer reasonable results. As is suggested by the differing values of ${x_2}^{\mathrm{P}}$, λ^{Q}, and ${u_0}^{\mathrm{P}}$ in Table 2, approximations to Eqn. (1) using $\hat{f}$ are generally less accurate than those using f. The comparison is given for the best results in the CDBM case, again obtained from $\mathbf{x}^{\mathbf{V}} = (\tilde{x_1^{\mathrm{P}}}, \tilde{x_2^{\mathrm{P}}})$. (Since CDBM results were shown to be preferable for f, OU metrics were not computed in the piecewise linear approximation). The accuracy of the closed-form expressions derived from the piecewise linear approximation $\hat{f}$ is sufficient to assist in our discussion of key quantitative and qualitative effects of parameters on the model's behavior, to which we now turn.

4 Discussion

4.1 Sensitivity analysis

To determine the sensitivity of reaction time statistics to perturbation around the standard parameter set, we compared the results of simulations with and without a perturbation (here 10% of the value of an individual parameter). For each RT statistic Y, a sensitivity measure S_i is assigned to every single-

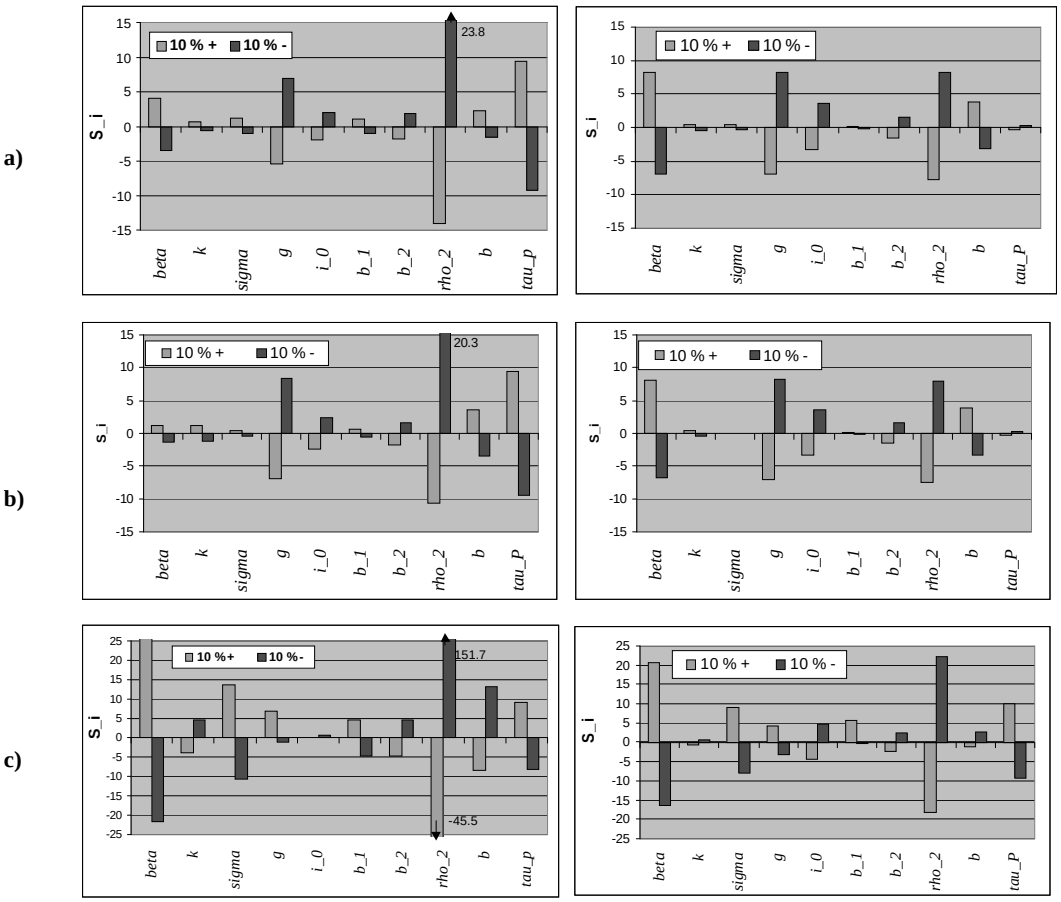


Figure 14: Results of the sensitivity analysis for 2-D numerical calculations (left column) and the piecewise-linear approximations using CDBM with $\mathbf{x}^{\mathbf{V}} = (\tilde{x_1^{\mathrm{P}}}, \tilde{x_2^{\mathrm{P}}})$ (right column). 2-D simulation results and the closed-form Eqn. (46) were numerically integrated to obtain corresponding means and variances of $R(t)$ for simulation and analytical results and the median for the simulation results; Eqn. (70) with V given by Eqn. (66) was used to obtain the median in the analytical approximation. a) S_i for the mean of $R(t)$; b) S_i for the median of $R(t)$; c) S_i for the variance of $R(t)$.

parameter perturbation $\Delta\phi_i$ from the standard parameter set as follows [12]:

$$S_i = \frac{Y_{\Delta\phi_i} - Y_{standard}}{Y_{standard}} \times 100\% \; . \tag{72}$$

Results of this sensitivity analysis are given in Fig. 14 for means and variances of reaction times. They were calculated using (left column) 2-D numerical solutions for the probability densities (Section 3.3) and (right column) the closed form expressions for $R(t)$ developed in Section 3.4.4 with $\mathbf{x^V} = (\tilde{x_1^{\mathrm{P}}}, \tilde{x_2^{\mathrm{P}}})$. These approximations reproduce many of the trends observed in the numerical solutions, suggesting the usefulness of the closed form expressions in determining the effects of parameters on reaction time distributions. The major differences between numerical results and analytical approximations are in the sensitivity of means and medians to τ_{P} and of all statistics to ρ_2. Analytical predictions for the former are incorrect in sign and magnitude; those for the latter, while qualitatively correct, significantly underestimate the sensitivities observed numerically.

4.2 Parameter study and relationships to cognitive control mechanisms

The methods developed above permit analysis of the influence of model parameters on simulated cognitive experiments. We now describe the dominant effects, focussing on i_0 and g: parameters whose adjustment may represent cognitive control mechanisms [6, 24, 8].

Reaction time means and medians are mainly determined by the magnitude of the trial vectorfield and the distance to the correct threshold from the post-preparatory starting point. The former effect may be observed in Figs. 14(a,b) as sensitivity to ρ_2 and b_2. Sensitivity to g demonstrates the latter effect, as the threshold value increases as g decreases, cf. Section 3.1. To the extent that they determine initial conditions for the trial phase, variations in i_0 also contribute to this effect.

Variances in reaction times are largely determined by the width of the distribution of initial conditions for trial trajectories. Fig. 2 shows the role of i_0 in determining whether preparatory cycle dynamics are in the monostable or bistable regimes, and hence the degree to which the relevant distribution can broaden during the preparatory period; however, this effect is not present in the parameter range of Fig. 14. Meanwhile, the value of τ_{P} determines how much of this broadening actually occurs, as is demonstrated in Fig. 14(c). With the notable exceptions of k and g, parameter changes that cause a decrease in mean and median reaction times also cause decreases in

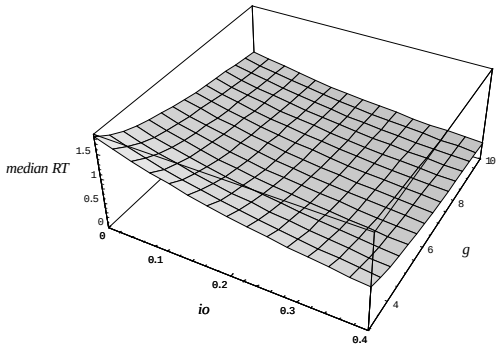


Figure 15: Variation in the median of $R^f(t)$ with g and i_0, with other parameters fixed at the standard values in the equal bias case and under the CDBM piecewise linear approximation with $\mathbf{x}^{\mathbf{V}} = (\tilde{x}_1^{\mathrm{P}}, \tilde{x}_2^{\mathrm{P}})$. From Eqns. (70), (67).

variances and vice-versa. This may be interpreted via the time-dependence of ν^2, as discussed in Section 3.5 in reference to Fig. 8.

Fig. 14(c) shows close agreement between 2-D numerical and 1-D piecewise linear predictions for the variation of RT medians with g and i_0. Under the latter approximation, Eqns. (70) and (67) may be used to evaluate effects of these parameters over a range of values. Fig. 15 shows the resulting variation in median RT; if unit biases remain equal, these equations would give similar plots for other parameters.

4.3 Relationships to Empirical Studies

The parameter dependencies revealed above are relevant to current cognitive psychology research on the anterior cingulate cortex (ACC) [6] and locus coeruleus (LC) [27] brain areas. It has been suggested that the ACC responds to high conflict signals (Eqn. (3)) by decreasing additive inputs i_0 to the decision units x_j. If the decrease is sufficient to break preparatory cycle bistability (cf. Fig. 2), it results in *controlled* response characteristics: specifically, longer reaction times and decreased error rates. Meanwhile, the LC's effects may be incorporated by modulation of gain g, which is thought to scale with LC-induced release of neuromodulators ([27], see also [24]), although this requires rapid gain changes during trials: an effect not considered here. Nevertheless, our analysis quantifies how the suppression of g and i_0 (in response to the presence of conflict) can result in more cautious response characteristics (Section 3.1 and Figs. 3(a), 15).

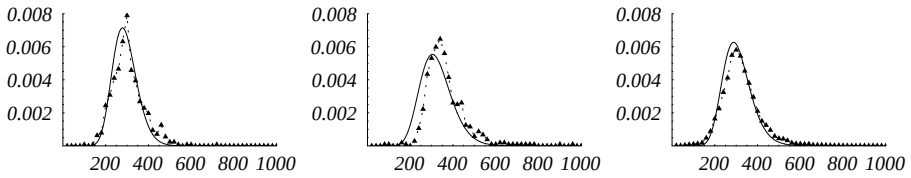


Figure 16: Comparison of $R(t)$ averaged over human subjects in behavioral trials (triangle datapoints) and $R(t)$ predicted by Eqn. (70) under the PW linear $\hat{f}$ and the standard parameter set excepting $\sigma = .24$ (solid line). Time units are msec. Left, for the AAAA stimulus history; center, for AAAR; right, averaged over all 16 stimulus histories.

To assess the modeling relevance of our standard parameter set, analytical predictions were compared with data from human subjects. Recent experimental results [19] for RSIs of 350 msec were chosen, since then RSI and median reaction times are approximately equal, as for the standard parameter set. Median reaction times predicted by Eqn. (70) for all sixteen stimulus histories were used to establish a linear relationship between model and experimental time units. Biases were predicted using, resp., shared and independent pattern detectors to register alternations and repetitions (Eqns. (4) and (5)); as per Eqns. (6-9), the detectors were functions of two prior stimuli. The fit was $t_e = t_m \times 107.0 + 206.2$ msec, where t_e and t_m are, resp., median experimental and model times. Comparisions of scaled and normalized $R(t)$'s from Eqn. (46) and from the experiments are shown in Fig. 16. To obtain appropriate RT variances, the noise level was increased by a factor of 1.7 to $\sigma \approx .27$. These results demonstrate that the model is capable of predicting RT distributions over the full range of stimulus histories, once the overall timescale has been fitted.

For the increased noise level used in Fig. 16, the maximum probability of threshold crossing during the preparatory phase is $1 - \int_0^\infty R(t)dt \approx .05$, an acceptably low value. This could be further reduced by choosing a lower σ and averaging over a distribution of 'subject' parameter values, as in Ratcliff et al. [22]. Alternatively, thresholds x_θ could be extended during the preparatory cycle to decrease $1 - \int_0^\infty R(t)dt$ to a negligible level. During the preparatory period, the new value of x_θ could be either retained or decreased dynamically (see [9] for a related approach); the latter option

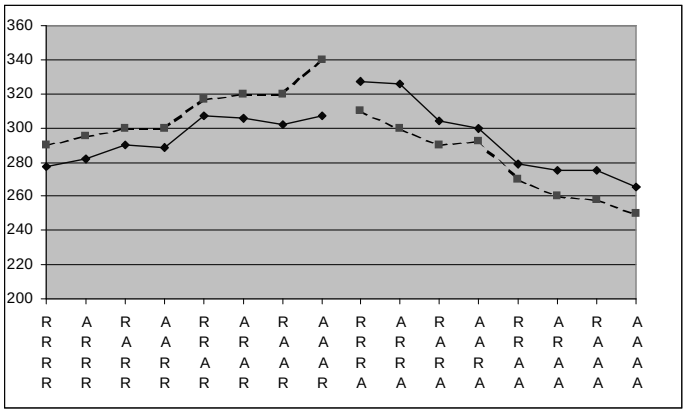


Figure 17: Comparison of RT medians for various stimulus histories. Dashed line, experimental data. Solid line, closed-form expressions for median of $R(t)$ (Eqn. (70) in Section 3.4.4 for $\mathbf{x}^{\mathbf{V}} = (\tilde{x_1^{\mathrm{P}}}, \tilde{x_2^{\mathrm{P}}})$). Time is in msec., with linear scaling and translation from model units as described in the text.

would induce a nonlinear time-dependence of the $v_{\theta j}$ (cf. Eqns. (39, 40)).

Next, the ability of the closed form expressions derived in Section 3.4.4 to reproduce experimentally observed trends in free-response reaction times [25] was tested. The standard parameter set was replaced by the average biases and other parameters used in producing Fig. 1 with the exception of $\tau_{\mathrm{P}} = 1$ (see section 2). To reflect this difference, median RTs from Eqn. (70) were compared with experimental results [25] for the 500msec RSI. Linear regression on all sixteen stimulus histories gave the relation $t_e = t_m \times 223.8 - 129.1$ msec. The resulting Fig. 17 shows that the analytical approximations capture much of the influence of stimulus history on RT medians. However, there is a clear disparity between experimental and model RT medians between the eight cases terminating in repeat (R) and the eight terminating in alternation (A): the model -R's are too fast and the -A's too slow. This might be corrected by redistribution of the repetition/alternation weights $\alpha_{A,R}$ in Eqns. (4, 5); c.f. [7].

Values of i_0, b_1, and b_2 used in Figs. 1 and 17 and in [7] differ significantly from the standard parameter set. Consequently, approximations developed for the standard parameter set were found to be generally less accurate in this regime. In a preliminary check for stimulus histories AAAA and AAAR,

setting $\mathbf{x}^{\mathbf{V}} = (x_{\theta 1}, x_\theta)$ and $\mathbf{x}^{\mathbf{V}} = (\tilde{x_1^{\mathrm{P}}}, \tilde{x_2^{\mathrm{P}}})$ gave preferable results for the L^1 and for the KL metrics of Section 3.5, resp., as for the standard parameter set. These parameters also provided an example of a regime in which the piecewise linear approximation to the logistic response is quite fragile. Specifically, the linear approximation exhibits bistability in several cases for which the logistic response function is monostable, resulting in qualitatively different dynamics. As γ is varied in the equal bias case, this occurs between the bifurcation values given by Eqns. (17) (logistic) and (51) (piecewise linear). For the standard parameter set, the corresponding parameter values are approximately $.02 < \gamma < .03$. Moreover, in the analog of the pitchfork bifurcation (Fig. 2) for the piecewise linear case, the asymmetric fixed points jump discontinuously off the diagonal as g passes through its bifurcation value. This extreme sensitivity with respect to g variation appears unrealistic.

5 Conclusions

This paper developed several methods for studying the relationship between neural network parameters and behavioral data produced by a simple model of a forced choice cognitive process. The model was described in Section 2, where preliminary analyses were performed and bifurcation diagrams drawn. A finite element solution to the corresponding Kolmogorov equation was established, and, in Section 3, a method due to Stone and Holmes [26] was extended to yield analytical estimates for decision probabilities and RT distributions. In Section 4, a sensitivity analysis of reaction times and error rates to perturbations around a standard set of parameters was performed, general conclusions regarding key parameter effects were drawn, and further comparisons with behavioral data were made.

In the analyses above, the control and bias parameters i_0 and b_j are regarded as constants while preparatory and trial dynamics are studied. In the Monte-Carlo simulations of [7], $i_0(n)$ and $b_j(n)$ are directly updated between trials. We also wish to study more general parameter update procedures in which i_0 and b_j depend continuously on time and updating (representing integration of previous experience) may occur *during* as well as between trials. Here, the effect of differing timescales will be crucial. In particular, the data of Soetens [25] displays significant modulation in RT median dependence on stimulus history as RSIs change, suggesting that subjects need sufficient time to "process" prior stimuli before new ones are presented. Further, the schemes we have examined here involve resetting the initial condition to the

origin for each trial. As mentioned in Section 2, one could alternatively continue evolving directly from the point where the trial trajectory crosses threshold. Work in progress includes studying the effects of these protocols on RT statistics.

In summary, the model developed above suggests that the dominant temporal dynamics of the neural network equations (1) can be represented as Ito diffusion on one-dimensional slow manifolds. Thus, as noted by Usher and McClelland [28], the drift-diffusion models of Ratcliff [21, 22] emerge naturally from this (small) connectionist model. Future research will include generalizing the methods developed here to account for more decision units and/or the possibility of multiple choices. In this case, a general connection matrix T_{ij} would be defined, and the model becomes

$$\dot{x_i} = -kx_i - \sum_j T_{ij} f(x_j) + i_0 + b_i + \rho_i + \eta_i \ ; \ i = 1 \ldots n \ . \tag{73}$$

A hypothesis we hope to test is that in certain cases the dominant temporal dynamics of Eqns. (73) can also be represented on slow manifolds, possibly of higher dimension.

Acknowledgements: This work was partially supported by DoE: DE-FG02-95ER25238 and NIMH: MH62196 (Cognitive And Neural Mechanisms of Conflict and Control, Silvio M. Conte Center). Eric Brown was supported under a National Science Foundation Graduate Fellowship. We thank Jonathan Cohen, Raymond Cho, and Leigh Nystrom for comments, corrections, and suggestions.

References

[1] D.J. Amit. *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge University Press, Cambridge, U.K., 1989.

[2] A.A. Andronov, E.A. Vitt, and S.E. Khaiken. *Theory of Oscillators*. Pergamon Press, Oxford, U.K., 1966. Reprinted by Dover Publications Inc, NY, 1987.

[3] L. Arnold. *Stochastic Differential Equations*. John Wiley, New York, 1974.

[4] G. Backstrom. *Fields of Physics by the Finite Element Method-In introduction.* Student Litteratur (ISBN 91-44-0655-1), Lund, Sweden, 1998.

[5] P. Bertelson. Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psych.*, 13:90–102, 1961.

[6] M. M. Botvinick, T.S. Braver, C.S. Carter, D.M. Barch, and J.D. Cohen. Conflict monitoring and cognitive control. *Psych. Rev.*, in press, 2001.

[7] R. Cho, L. Nystrom, E. Brown, A. Jones, T. Braver, P. Holmes, and J. Cohen. Mechanisms underlying performance dependencies in a two-alternative forced choice task. In preparation, 2001.

[8] J.D. Cohen and D. Servan-Schreiber. Context, cortex and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psych. Rev.*, 99:45–77, 1992.

[9] J.D. Cohen et al. Mechanisms of spatial attention: The relation of macrostructure to microstructure in parietal neglect. *J. Cog. Neurosci.*, 6(4):377–387, 1994.

[10] B.D. Coller. *Suppression of Heteroclinic Bursts in Boundary Layer Models*. PhD thesis, Cornell University, 1995.

[11] T. Cover. *Elements of Information Theory*. Wiley, New York, 1991.

[12] R. Dickenson and R. Gelinas. Sensitivity analysis of ordinary differential equation systems – A direct method. *J. Comp. Phys.*, 21 (2), 1976.

[13] C. W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*. Springer-Verlag, New York, 1983.

[14] J. Guckenheimer and P.J. Holmes. *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*. Springer-Verlag, New York, 1983.

[15] P. Holmes, J.L. Lumley, and G. Berkooz. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press, Cambridge, U.K., 1996.

[16] J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA*, 79:2554–2558, 1982.

[17] J.J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA*, 82:3088–3092, 1984.

[18] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, 1991.

[19] L. Nystrom, R. Cho, and J. Cohen. Princeton Center for the Study of Mind, Brain, and Behavior Technical Report, in preparation, 2001.

[20] B. Oksendal. *Stochastic Differential Equations*. Springer-Verlag, New York, 1995.

[21] R. Ratcliff. A theory of memory retrieval. *Psych. Rev.*, 85 (2):59–108, 1978.

[22] R. Ratcliff, T. Van Zandt, and G. McKoon. Connectionist and diffusion models of reaction time. *Psych. Rev.*, 106 (2):261–300, 1999.

[23] D.E. Rumelhart and J.L. McClelland. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press, Cambridge, MA., 1986.

[24] D. Servan-Schreiber, H. Printz, and J.D. Cohen. A network model of catecholamine effects: Gain, signal-to-noise ratio, and behavior. *Science*, 249:892–895, 1990.

[25] E. Soetens et al. Expectancy or automatic facilitation? Separating sequential effects in two-choice reaction times. *J. Experimental Psych: Human Perception and Performance.*, 11:598–616, 1985.

[26] E. Stone and P. Holmes. Random perturbations of heteroclinic cycles. *SIAM J. on Appl. Math.*, 50(3):726–43, 1990.

[27] M. Usher, J.D. Cohen, J. Rajkowsky, P. Kubiak, and G. Aston-Jones. The role of locus coeruleus in the regulation of cognitive performance. *Science*, 283:549–554, 1999.

[28] M. Usher and J.L. McClelland. On the time course of perceptual choice: The leaky competing accumulator model. *Psych. Rev.*, 2001. In press.