International Journal of Bifurcation and Chaos, Vol. 15, No. 3 (2005) 803–826
© World Scientific Publishing Company

SIMPLE NEURAL NETWORKS THAT OPTIMIZE DECISIONS

ERIC BROWN*, JUAN GAO†, PHILIP HOLMES*,†, RAFAL BOGACZ*,‡,
MARK GILZENRAT‡, and JONATHAN D. COHEN‡

*Program in Applied and Computational Mathematics,
†Department of Mechanical and Aerospace Engineering,
‡Department of Psychology,
Princeton University, Princeton, NJ 08544, USA

Received April 2, 2004; Revised July 7, 2004

We review simple connectionist and firing rate models for mutually inhibiting pools of neurons that discriminate between pairs of stimuli. Both are two-dimensional nonlinear stochastic ordinary differential equations, and although they differ in how inputs and stimuli enter, we show that they are equivalent under state variable and parameter coordinate changes. A key parameter is gain: the maximum slope of the sigmoidal activation function. We develop piecewise-linear and purely linear models, and one-dimensional reductions to Ornstein–Uhlenbeck processes that can be viewed as linear filters, and show that reaction time and error rate statistics are well approximated by these simpler models. We then pose and solve the optimal gain problem for the Ornstein–Uhlenbeck processes, finding explicit gain schedules that minimize error rates for time-varying stimuli. We relate these to time courses of norepinephrine release in cortical areas, and argue that transient firing rate changes in the brainstem nucleus locus coeruleus may be responsible for approximate gain optimization.

Keywords: Gain; neural network model; decision task; stochastic differential equation; reaction time; optimal speed and accuracy; matched filter; locus coeruleus.

1. Introduction

The psychological and neural bases of decision making are active areas of inquiry in cognitive science [Schall, 2001; Gold & Shadlen, 2001; Schall *et al.*, 2002; Gold & Shadlen, 2002; Shadlen & Newsome, 2001; Platt & Glimcher, 1999; Stone, 1960; Laming, 1968; Ratcliff, 1978; Ratcliff *et al.*, 1999; Usher & McClelland, 2001; Roitman & Shadlen, 2002; Wang, 2002]. There is a wealth of data on simple decision tasks which require discrimination among alternative stimuli as quickly and accurately as possible. Typically, this discriminatory process has been modeled as a competition among different neural populations, each representing alternate interpretations of the current stimulus [Cohen *et al.*, 1990; Usher & McClelland, 2001]. Recent direct recordings in visual and motor areas of monkeys performing sensory discrimination tasks support this interpretation by revealing that, following training, certain "decision" neurons become selective for different stimulus alternatives, and upon presentation of the relevant stimulus their firing rates gradually increase accordingly; when these rates cross thresholds, the corresponding behavioral response is initiated (e.g. [Schall, 2001; Gold & Shadlen, 2001; Schall *et al.*, 2002; Roitman & Shadlen, 2002; Gold & Shadlen, 2002]). This neural evidence adds to behavioral evidence noted below, suggesting that decisions are made by comparing integrated "weights of evidence", encoded by the firing rates of neural groups. Here, we explore

the computational mechanisms required to optimize such a process.

The stimuli relevant to making a decision are often not static: their saliences may change over time. In the simplest case, a change occurs only at the moment when the stimulus itself appears. This is typically modeled in simulations of decision tasks (e.g. in [Cohen & Huston, 1994; Brown & Holmes, 2001; Cho *et al.*, 2002], cf. [Laming, 1968]) by dividing the task into two distinct periods: a preparatory period, in which no stimulus is present, and a trial period, in which a stimulus of constant discriminability is presented. Alternatively, stimulus discriminability may change in a stepwise manner or vary continuously.

The following specific example motivates our analysis of two specific cases in Sec. 2.5. In the "moving dots" paradigm of the two alternative forced choice task [Britten *et al.*, 1993; Shadlen & Newsome, 2001; Gold & Shadlen, 2002] a display of moving dots is presented, and the subject must indicate whether a majority of dots is moving to the right or the left. In the simplest case, the subject focuses on a neutral fixation point during the preparatory period, after which the dots appear, with a certain "coherent" fraction moving either left or right, and the rest moving randomly. A variant is obtained by showing a zero coherence display of dots during the preparatory period, and suddenly increasing coherence to a fixed value.

Even if external stimuli have constant strengths, their representations in neural populations that decide between alternative hypotheses may *gradually* rise, due to accumulating activity in input layers, fluctuations in attention, or both [Mozer, 1988; Cohen *et al.*, 1992; Usher *et al.*, 1999; Gilzenrat *et al.*, 2002]. Another possible source of time varying salience is the increasing noise levels that may accompany higher firing rates. A richer situation, in which the stimulus salience increases and decreases over time, is explored in [Huk *et al.*, 2002]. A focus of the present paper is how stimuli with time-dependent salience can be *optimally* processed in simple neurally-based models of decision networks. We study the reduction of such networks to linearized, one-dimensional approximations (cf. [Usher & McClelland, 2001; Brown & Holmes, 2001; Bogacz *et al.*, 2004]) for which optimality conditions can be fully characterized, and identify two distinct mechanisms, one involving intrinsic properties of decision networks and the other involving external modulation, that can implement optimal processing of time-varying stimuli.

Optimality principles have found wide application in psychology and neuroscience (e.g. [Bialek *et al.*, 1991; Anderson, 1990; Fairhall *et al.*, 2001]). In particular, Stone [1960] applied the optimal Sequential Probability Ratio Test (SPRT) to model behavioral data in a two-alternative forced choice task. This was followed by the extensive work of Laming [1968]. The SPRT computes time-dependent likelihood ratios between the probabilities of two competing hypotheses, a procedure equivalent to the signal processing strategy that maximizes signal-to-noise ratio in the difference between two incoming stimuli. For stimuli with constant signal-to-noise ratios, the SPRT is equivalent, in an appropriate continuum limit, to the constant-drift diffusion model, which has been shown by Ratcliff and others to fit a wide variety of behavioral data (see [Ratcliff, 1978; Ratcliff *et al.*, 1999] and references therein) and also to describe the dynamics of neural firing rates in sensori-motor brain areas [Schall *et al.*, 2002; Gold & Shadlen, 2002], cf. [Smith & Ratcliff, 2004]. Specifically, in [Gold & Shadlen, 2002], the notion of reward rate is introduced for the constant-drift diffusion model, and [Bogacz *et al.*, 2004] shows that higher performing subjects do optimize this quantity in a specific behavioral task. However, although [Laming, 1968] does allow for accumulation of noise to have occurred before stimulus presentation (see Laming's Appendix A7), in all these studies the decision process is modeled only *after* presentation of a stimulus having constant signal-to-noise ratio; furthermore, the parameters describing processing of incoming information are not explicitly allowed to vary in time.

In this paper we show how models of mutually inhibiting neural populations can make nearly optimal decisions about the identity of *time-varying* stimuli. This is accomplished via dynamical adjustments in an *effective* gain parameter for the linearized population dynamics. The gain determines the sensitivity of (equilibrium) population firing rates to changes in averaged input currents to the population, and the word "effective" is used here because these changes can result either from transient variations in the gain parameter describing this sensitivity or directly from the nonlinearities of neural input–output functions. There is much current research into neural mechanisms for the modulation of gain in neural populations,

identifying such factors as levels of norepinephrine [Usher *et al.*, 1999] and the strength of fluctuations in individual neurons comprising the population (e.g. [Chance *et al.*, 2002; Amit & Tsodyks, 1991; Brunel *et al.*, 2001]). In particular, Shin *et al.* [1999] proposes a mechanism in which frequency-current curves of individual neurons adapt to match operating ranges to neural inputs, via intracellular calcium signals. This may be viewed as a biophysical implementation of the earlier "automatic gain control" (see Eq. (9) of [Grossberg, 1988] and references therein), which is implemented via multiplicative "shunting" terms in neural network models and also keeps neural units in the sensitive regimes of their input–output functions. Gain plays a different role in the present paper: we identify, for three different models, the distinct time-dependent (effective) gain schedules which implement optimal processing strategies for time-dependent signals. These provide predictions for gain manipulations that diverse neural mechanisms may implement to improve task performance.

The balance of the paper proceeds as follows. In Sec. 2 we introduce the forced and free response decision tasks, and three types of stochastic differential equation (SDE) models for these tasks. We show that two of these are related via a coordinate transformation, and discuss linearized and one-dimensional reductions of them, exploring the accuracy of these reductions in two rather general cases. In the following Sec. 3, we compute time-dependent values of gain that optimize signal processing in the one-dimensional models. This involves calculating gain functions that enable them to implement the classical signal processing notion of matched filters. Section 4 interprets these results in terms of cortical norepinephrine (NE) release mediated by the brainstem nucleus locus coeruleus (LC), showing that LC and NE dynamics indeed appear to approximate optimal time courses. Section 5 concludes the paper with a brief discussion.

Although we only consider simple models of a prototypical cognitive task, we believe that this paper is appropriate for a volume celebrating the centenary of John von Neumann's birth. Early in 1956 von Neumann was working on a manuscript in preparation for the Silliman memorial lectures at Yale, which he had been invited to deliver that Spring. Unfortunately, his final illness intervened and he entered the Walter Reed Hospital in April, where he remained until his death in February 1957. The lectures were never given, but his remarkable book, *The Computer and the Brain* [von Neumann, 1958], remains among his final work. In it, he makes elegant and simple estimates of human neural computational capacity based on notions drawn from the theory of analog and digital automata (which he had largely developed), and from information theory. Although neuronal spikes appear as 1's (and their absence as 0's), he argues that neural computation is necessarily inaccurate and noisy, and hence must be "statistical" rather than "digital." He points out that firing rates in sensory neurons tend to be monotone functions of stimulus strength and, as an early proponent of rate coding, he can be seen as pioneering the class of firing rate models treated here.

2. Models of Decision Tasks

2.1. *Decision tasks: The forced and free response protocols*

We consider two distinct tasks, both widely used in cognitive neuroscience, in each of which a decision maker must discriminate between two alternatives, henceforth denoted "1" and "2". The sensory information itself, as well as its neural representation, is assumed to be noisy, so that discrimination errors occur. The first task is the *forced-response* paradigm, in which subjects must respond at a fixed time T following stimulus onset with their best estimate of which alternative (1 or 2) was presented. Performance on this task is measured by the error rate, or one minus the fraction of correct responses. We will also refer to this as the *interrogation protocol*, noting that it is distinct from deadlining (not considered further here), in which subjects are apprised in advance of a fixed, maximal time *before* which all responses must be made.

In the second, *free-response* paradigm, decisions are not demanded at a preset time, but are given when the subject feels that sufficient evidence in favor of one alternative has accumulated. Since the sensory evidence is noisy, response times vary from trial to trial and performance under the free-response condition is characterized by both reaction times and error rates. Here, optimality requires an appropriate balance of speed and accuracy [Wickelgren, 1977; Gold & Shadlen, 2002; Bogacz *et al.*, 2004].

Following [Usher & McClelland, 2001] and others, we shall model both these tasks by a pair of competing (mutually inhibitory) neural

populations, each of which is selectively responsive to sensory input corresponding to one of the two alternatives. In the forced-response protocol, the neural population with the highest firing rate at time T determines the decision. For free responses, the first of the two populations to cross a firing rate threshold establishes the choice. We do not address the (interesting) question of how thresholds are set or threshold crossings are detected.

2.2. *Two-dimensional nonlinear models and the neural gain parameter*

In this section we consider the dynamics of two mutually inhibiting neural populations, each of which receives noisy sensory input from components of the stimulus representing one of the alternatives. We describe two models for such populations, both in wide use, and both in the form of systems of stochastic ordinary differential equations (SDEs) [Arnold, 1974].

The first of these, the leaky integrator *connectionist* model [McClelland, 1979; Usher & McClelland, 2001], is:

$$\tau_c \frac{dx_1}{dt} = -x_1 - \beta f_{g(t)}(x_2) + a_1(t) + \frac{c(t)}{\sqrt{2}}\eta_t^1, \quad (1)$$

$$\tau_c \frac{dx_2}{dt} = -x_2 - \beta f_{g(t)}(x_1) + a_2(t) + \frac{c(t)}{\sqrt{2}}\eta_t^2, \quad (2)$$

where the state variables $x_j(t)$ denote the mean input currents to cell bodies of the jth neural population, the integration implicit in the differential equations modeling temporal summation of dendritic synaptic inputs ([Grossberg, 1988] and references therein). Additionally, the parameter β sets the strength of mutual inhibition via population firing rates $f_{g(t)}(x_j(t))$, where $f_{g(t)}(\cdot)$ is the sigmoidal "activation" (or "frequency-current" or neural "input–output") function to be described shortly. The stimulus signal received by each population is $a_j(t)$, and the noise terms polluting this signal are $c(t)\eta_t^j$, where $c(t)$ sets r.m.s. noise strength and the η_t^j are (independent) white noise processes with variance $\mathrm{E}(\eta_t^j - \eta_{t'}^j)^2 = \delta(t-t')$. The time constant τ_c reflects the rate at which neural activities decay in the absence of inputs and respond to input changes. Under the free-response paradigm a decision is made and the response initiated when the firing rate $f_{g(t)}(x_j)$ of either population first exceeds a preset threshold θ_j; it is normally assumed that $\theta_1 = \theta_2 = \theta$. For the interrogation protocol, the population with greatest activity (and also firing rate) at time T determines the decision. We also assume that activities decay to zero after response and prior to the next trial, so that the initial conditions for (1)–(2) are $x_j(0) = 0$.

The subscript in $f_{g(t)}(\cdot)$ indicates dependence on the time-varying gain, or sensitivity, $g(t)$ of the neural populations: gain sets the slope of the activation function. For example, the logistic function

$$f_{g(t)}(x) = \frac{1}{1+\exp(-4g(t)(x-b))} = \frac{1}{2}[1+\tanh(2g(t)(x-b))] \quad (3)$$

has maximal slope $g(t)$ (see Fig. 1, left). While this specific form is not required for the results derived below, we do assume that f_g takes its time-dependent maximal slope $g(t)$ at some time-independent point, as for (3).

As already mentioned, the connectionist model describes the time evolution of current inputs. A second model is derived in [Wilson & Cowan, 1972], cf. [Hopfield, 1984; Abbott, 1991; Gerstner & Kistler, 2002], in which the firing rates of neural populations are themselves integrated over time. First we give the linearized version of this *firing rate* model:

$$\tau_c \frac{dy_1}{dt} = -y_1 + f^l_{g(t)}\left(-\beta y_2 + a_1(t) + \frac{c(t)}{\sqrt{2}}\eta_t^1\right), \quad (4)$$

$$\tau_c \frac{dy_2}{dt} = -y_2 + f^l_{g(t)}\left(-\beta y_1 + a_2(t) + \frac{c(t)}{\sqrt{2}}\eta_t^2\right). \quad (5)$$

Here, the y_j are the firing rates of population j and other terms are as above. The linear function

$$f^l_{g(t)}(x) = \frac{1}{2} + g(t)\,(x-b)\ , \quad (6)$$

derives from replacing the logistic (or any similar monotonic) function by the linear approximation $f^l_{g(t)}(\cdot)$ around its point of maximal slope. Note that

the firing rate y_j of the jth population approaches an equilibrium set by the input currents to this population, passed through the (linearized) frequency-current function. This model must be reformulated to allow for nonlinear functions $f_{g(t)}$, because white noise does not make sense as an argument in such a function, cf. [Gardiner, 1985]. In particular, we assume that, as in (4)–(5), the strength of firing rate fluctuations in response to noise in inputs scales with $g(t)$ (i.e. with the maximal sensitivity of firing rates to the deterministic component of the input). This yields

$$\tau_c \frac{dy_1}{dt} = -y_1 + f_{g(t)}(-\beta y_2 + a_1(t)) + g(t)\frac{c(t)}{\sqrt{2}}\eta_t^1, \tag{7}$$

$$\tau_c \frac{dy_2}{dt} = -y_2 + f_{g(t)}(-\beta y_1 + a_2(t)) + g(t)\frac{c(t)}{\sqrt{2}}\eta_t^2, \tag{8}$$

which is valid for all $f(\cdot)$ and reduces to the form (4)–(5) for linear $f(\cdot)$. Note that the firing rate model (7)–(8) is a standard two-unit recurrent neural network with additive noise [Hertz *et al.*, 1991]. As above, we take initial conditions $y_j(0) = 0$, and note that threshold-crossing in the free-response case is detected directly via $y_j = \theta_j$.

For the questions of optimal stimulus processing addressed here, the most important distinction between the connectionist (1)–(2) and firing rate (4–5)–(7–8) models is whether the inputs $a_j(t) + c(t)/\sqrt{2}\eta_t^j$ enter as separate additive terms, as in the former, or as arguments to the activation function $f_g(t)$, as in the latter. As explained at the end of Sec. 3, this determines whether changes in gain directly adjust the sensitivity of neural units to all inputs or just to feedback from the competing unit, and it results in qualitatively different predictions for optimal gain schedules in the two models. While we expect that future work on low-dimensional descriptions of the population dynamics of spiking neurons (extending, e.g. [Brunel *et al.*, 2001; Wang, 2002; Omurtag *et al.*, 2000; Shelley & McLaughlin, 2002; Ermentrout, 1994] to include neurotransmitter effects) will result in more refined models, here we study the "simple" connectionist and firing rate descriptions. Throughout, we use variables x_j in referring to the former and y_j to the latter.

2.3. *Equivalence of the firing rate and connectionist models*

We now show that the firing rate and connectionist models are equivalent under a (generally time-dependent) coordinate change and corresponding adjustment of parameters, initial conditions, and thresholds. Specifically, for any activation function that is odd around some input value, such as (3), (7)–(8) can be written in the form (1)–(2). Hence, for every parameterization of the firing rate model, there is a connectionist model that produces identical trajectories as well as error rate and reaction time statistics, and vice-versa. This shows that the two models are effectively equivalent, up to parameterization. However, in Sec. 3 below we demonstrate that, because of the different ways that gain $g(t)$ enters them, their optimal gain trajectories differ significantly.

Starting with Eqs. (7)–(8), we extend the S–Σ exchange transformation of Grossberg [1988] to define the new coordinates

$$\tilde{y}_1 = 2b + \beta y_1 - a_2, \quad \tilde{y}_2 = 2b + \beta y_2 - a_1\,, \tag{9}$$

so that $-\beta y_1 + a_2 = -\tilde{y}_1 + 2b$ and $-\beta y_2 + a_1 = -\tilde{y}_2 + 2b$. In terms of these (7)–(8) become

$$\tau_c \frac{d\tilde{y}_1}{dt} = \beta\left[-\frac{1}{\beta}(a_2 + \tilde{y}_1 - 2b) + f_{g(t)}(-\tilde{y}_2 + 2b) + g(t)\frac{c(t)}{\sqrt{2}}\eta_t^1\right] - \frac{da_2}{dt}\,, \tag{10}$$

$$\tau_c \frac{d\tilde{y}_2}{dt} = \beta\left[-\frac{1}{\beta}(a_1 + \tilde{y}_2 - 2b) + f_{g(t)}(-\tilde{y}_1 + 2b) + g(t)\frac{c(t)}{\sqrt{2}}\eta_t^2\right] - \frac{da_1}{dt}, \tag{11}$$

and using the following property of the logistic activation function (3):

$$\begin{aligned} f_{g(t)}(-\xi + 2b) &= \frac{1}{2}[1 + \tanh(2g(t)[-\xi + 2b - b])] \\ &= \frac{1}{2}[1 + \tanh(-2g(t)[\xi - b])] \\ &= \frac{1}{2}[1 - \tanh(2g(t)[\xi - b])] \\ &= 1 - \frac{1}{2}[1 + \tanh(2g(t)[\xi - b])] \\ &= 1 - f_{g(t)}(\xi), \end{aligned} \tag{12}$$

(10)–(11) become

$$\tau_c \frac{d\tilde{y}_1}{dt} = -\tilde{y}_1 - \beta f_{g(t)}(\tilde{y}_2) - a_2 - \dot{a}_2 + 2b + \beta + \beta g(t)\frac{c(t)}{\sqrt{2}}\eta_t^1,$$

$$\tau_c \frac{d\tilde{y}_2}{dt} = -\tilde{y}_2 - \beta f_{g(t)}(\tilde{y}_1) - a_1 - \dot{a}_1 + 2b + \beta + \beta g(t)\frac{c(t)}{\sqrt{2}}\eta_t^2.$$

This SDE has the same form as (1)–(2) with parameters mapped as follows:

$$\begin{aligned} a_1 &\mapsto 2b + \beta - a_2 - \dot{a}_2, \\ a_2 &\mapsto 2b + \beta - a_1 - \dot{a}_1. \end{aligned} \tag{13}$$

The firing rate model (7)–(8) therefore produces identical statistics to the connectionist model (1)–(2) with appropriately remapped parameters and state variables. Note that thresholds and initial conditions for the firing rate variables y_1, y_2 must be transformed under (9) to apply to the equivalent connectionist model, that a_1 and a_2 are interchanged in the inputs, and that the noise terms are multiplied by gain $g(t)$.

2.4. *Piecewise-linear approximations*

As in [Usher & McClelland, 2001; Brown & Holmes, 2001], Eq. (3) may be approximated by a piecewise-linear function:

$$\begin{aligned} f_{g(t)}(\xi) &\approx f^{pw}_{g(t)}(\xi) \\ &= \begin{cases} 0 & \text{for } \xi \in \left(-\infty, b - \dfrac{1}{2g}\right] \\ \dfrac{1}{2} + g(t)(\xi - b) & \text{for } \xi \in \left[b - \dfrac{1}{2g}, b + \dfrac{1}{2g}\right], \\ 1 & \text{for } \xi \in \left[b + \dfrac{1}{2g}, \infty\right) \end{cases} \end{aligned} \tag{14}$$

as illustrated in Fig. 1. Note that our choice to set the slope of $f^{pw}_{g(t)}$ in its central domain equal to the maximal slope $g(t)$ of the nonlinear function $f_{g(t)}$ does not minimize the distance between the two functions in the L^∞ or L^2 norms. The best L^∞ match is obtained by setting the maximal slope of $f^{pw}_{g(t)}$ equal to $0.71g(t)$, and in L^2 by a $g(t)$-dependent value ranging between $0.72g(t)$ and $0.76g(t)$ (for $g(t)$ between 0.25 and 3). However, all

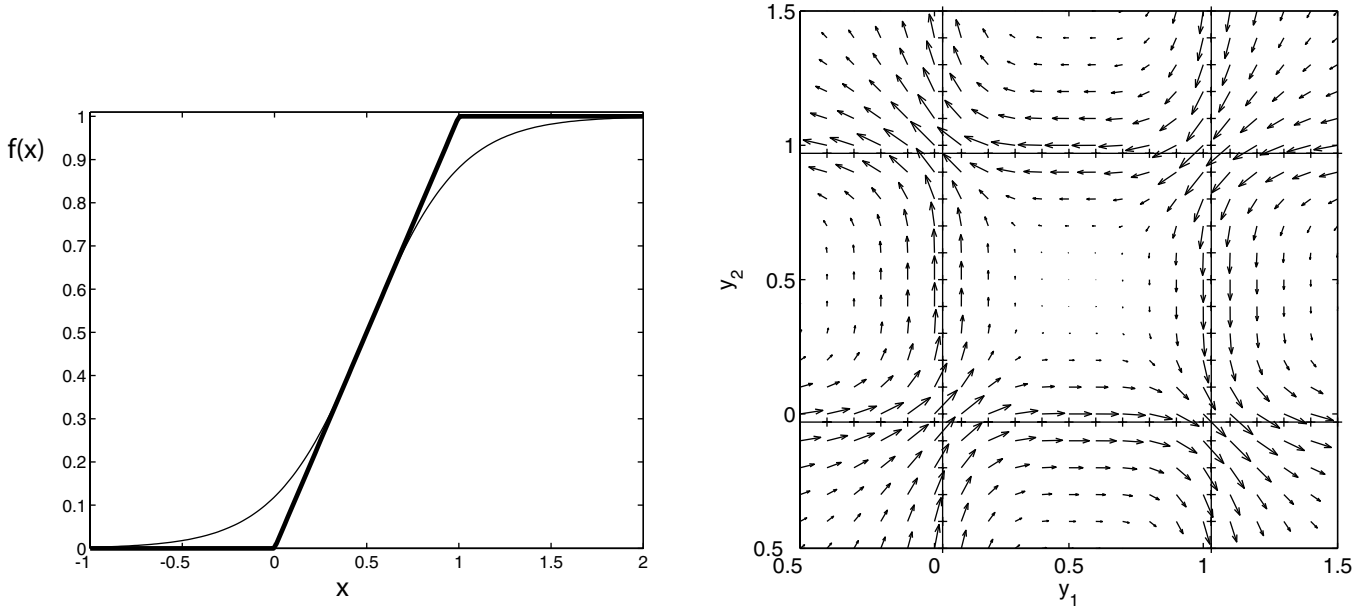


Fig. 1. (Left) Comparison of logistic and piecewise-linear activation functions; $g = 1, b = 0.5$. (Right) Comparison of logistic and piecewise-linear vectorfields $F(y_1, y_2)$ and $F^{pw}(y_1, y_2)$ for the piecewise-linear firing rate model (15)–(16): the difference $F(y_1, y_2) - F_{pw}(y_1, y_2)$ is plotted. Also shown for reference are the nine phase space tiles described in Fig. 2. Here additionally $\tau_c = 1$, $\beta = 1$, $a_1 = 1.03$, $a_2 = 0.97$.

these choices result in similar error rate and reaction time statistics, and we use (14) in what follows.

For ease of reference, we rewrite Eqs. (7)–(8) following piecewise linearization:

$$\tau_c \frac{dy_1}{dt} = -y_1 + f^{pw}_{g(t)}(-\beta y_2 + a_1(t)) + g(t)\frac{c(t)}{\sqrt{2}}\eta^1_t, \tag{15}$$

$$\tau_c \frac{dy_2}{dt} = -y_2 + f^{pw}_{g(t)}(-\beta y_1 + a_2(t)) + g(t)\frac{c(t)}{\sqrt{2}}\eta^2_t. \tag{16}$$

The difference between the vectorfield of the fully nonlinear model (7)–(8) and that of (15)–(16) is illustrated in Fig. 1 (right) for a specific parameter choice. In Sec. 2.6 below, we shall explicitly compare reaction times and error rates predicted by these two models.

The (y_1, y_2) phase space of the piecewise-linear firing rate model (and of the analogous connectionist model) is tiled by nine regions divided by pairs of horizontal and vertical lines at the break points of f^{pw}_g, each having a distinct linear vectorfield: see Fig. 2. In the following section, we will describe two cases in which this tiled structure can be used to reduce Eqs. (7)–(8) to a one-dimensional system.

2.5. *Representing decision dynamics in one dimension*

As discussed above and in [Usher & McClelland, 2001], in the forced response protocol, the choice $j = 1$ or 2 is made according to which of the two neural populations has the greatest activity or firing rate at interrogation time T. Therefore, knowledge of the difference

$$\begin{aligned} y(T) &\triangleq y_1(T) - y_2(T) \quad \text{or} \\ x(T) &\triangleq x_1(T) - x_2(T) \end{aligned} \tag{17}$$

determines the outcome and reduction of the original two-dimensional problem to a single variable does not *inherently* imply any loss in accuracy. For example, if the difference in firing rates is described by a time-dependent probability density $p(y, t)$ (whose distribution represents variability across behavioral trials), then the error rate at

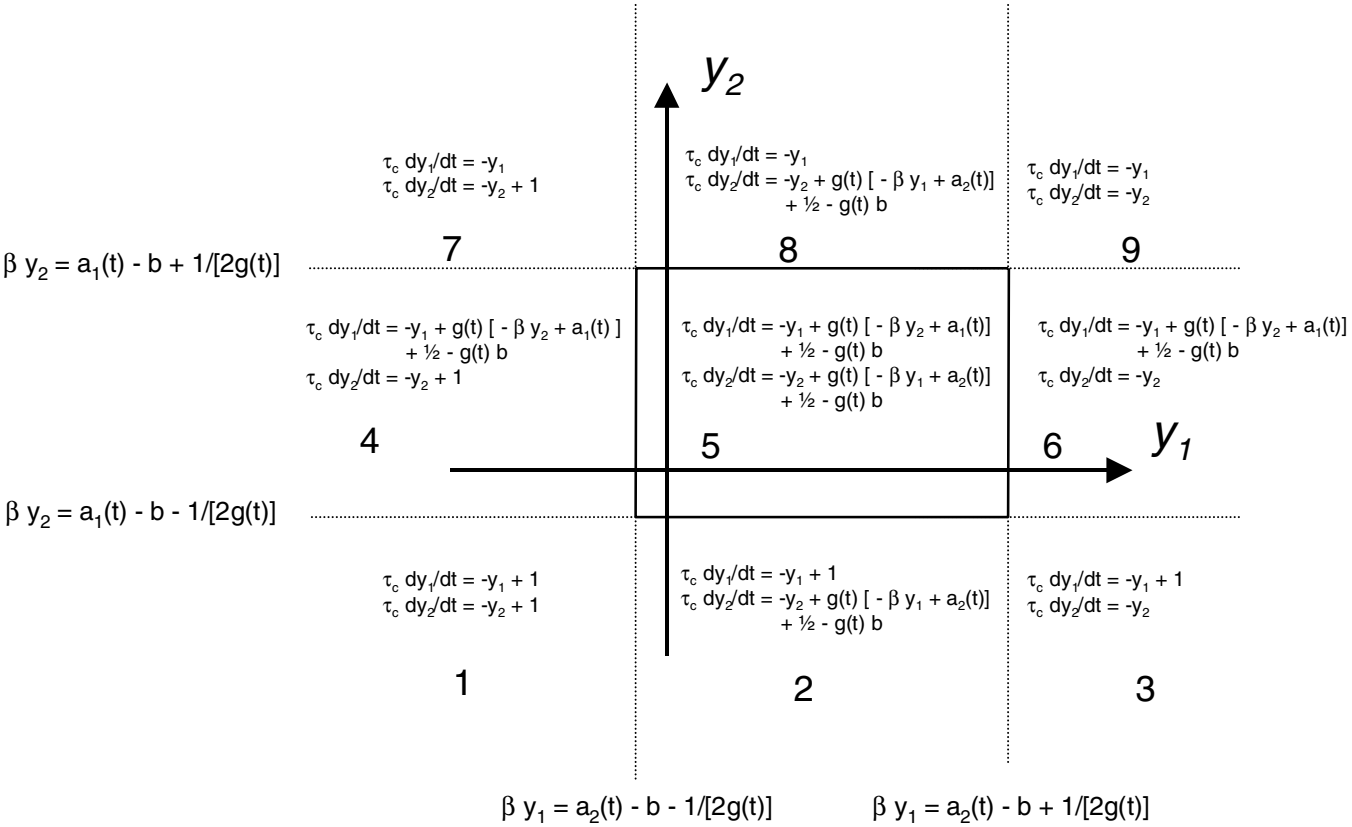


Fig. 2. The piecewise-linear vectorfield of the firing rate model (15)–(16). The central tile is surrounded by a solid box.

interrogation time T is

$$ER = \int_0^\infty p(y,T)dy \tag{18}$$

if alternative 2 was presented (i.e. if $a_2 > a_1$ for $t > t_s$), and

$$ER = \int_{-\infty}^0 p(y,T)dy \tag{19}$$

if alternative 1 was presented. Similar conclusions hold for the connectionist model.

For the free choice protocol the situation is more subtle. The single variable x or y is sufficient to characterize the decision only if the probability density of solutions to (7)–(8) or (1)–(2) has approximately collapsed along a one-dimensional "decision manifold" $\mathcal{M}$ by the time the threshold is crossed; see Fig. 3. In this sketch, the decision manifold, parameterized by y, is the unstable, center or weak stable manifold [Guckenheimer & Holmes, 1983] of the indicated fixed point, which, for the linearized system, coincides with its eigenspace.

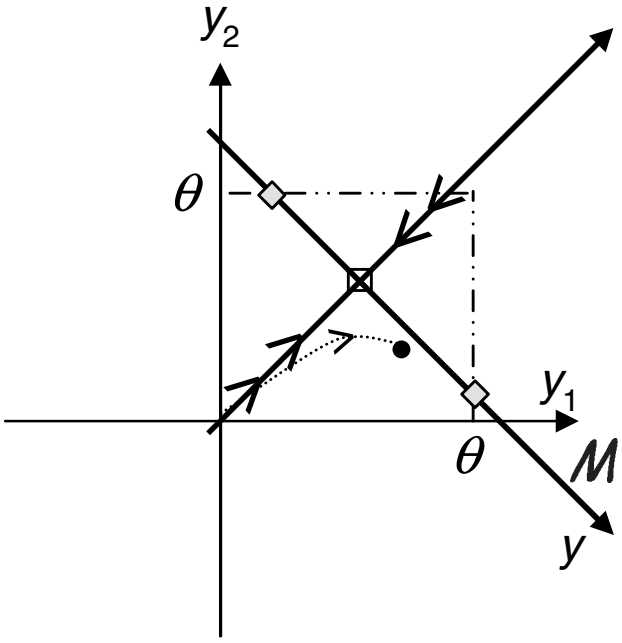


Fig. 3. Reduction to one dimension. The coordinate y (or x) of Eq. (17) parameterizes the decision manifold $\mathcal{M}$ (see text): the invariant manifold containing the fixed point indicated by the square. In the free response protocol, collapse of noisy solutions along $\mathcal{M}$ is required for accurate description in one dimension (cf. Figs. 4 and 5 (right)) so that sample paths (dotted line and point) cross thresholds arbitrarily close to the intersections of $\mathcal{M}$ with the thresholds $y_j = \theta$. This is not required for the forced response/interrogation protocol, in which the probability density $p(y,t)$ is simply cut along $y_1 = y_2$ at $t = T$.

The existence of center manifolds $\mathcal{M}$ for SDEs with additive noise, such as those considered here, has been proven rather generally: see [Boxler, 1991] and [Arnold, 1998, Chap. 7]; also [Knobloch & Wiesenfeld, 1983] for an early analysis and explicit examples. However, here we consider only the fully linear and piecewise linear systems, for which the "diagonal" coordinates $y = y_1 - y_2$, $\tilde{y} = y_1 + y_2$ and assumption of independent white noise processes decouple the components of (7)–(8) (and analogously of (1)–(2)) [Bogacz *et al.*, 2004], and so we do not need the full power of these results.

For collapse to $\mathcal{M}$ to occur, the eigenvalue characterizing dynamics normal to the manifold must be sufficiently negative compared with the other eigenvalue and the noise strength c, so that the joint probability density $p(y_1, y_2, t)$ rapidly concentrates near $\mathcal{M}$ and a substantial majority of sample paths crosses the thresholds $x_j = \theta$ (or $y_j = \theta$) near their intersections with $\mathcal{M}$ [Usher & McClelland, 2001; Brown & Holmes, 2001; Bogacz *et al.*, 2004]. These requirements are met by two distinct parameter sets to be introduced below, and in Sec. 2.6 we compare the resulting reaction times and error rates determined from one-dimensional reductions with those of the original two-dimensional models.

2.5.1. *Dimension reduction and transient gain in two simple cases*

In two cases, a simple equation for the evolution of $x(t)$ or $y(t)$ may be derived. These cases are characterized by a dominant proportion of solutions to (15)–(16) (i.e. for "most" realizations of the noise processes $\eta_j(t)$) (i) being confined to a single tile for the duration of the decision process or (ii) "jumping" together between tiles. The first of these situations occurs for Case 1 parameter sets, in which, for example, the onset of salience (i.e. $a_1 \neq a_2$) in input currents is accompanied by large transients in the magnitude of these inputs. The second Case 2 occurs for stimuli in which salience appears without such transients in magnitude. We now consider these cases in detail for the firing rate model.

Case 1. Trajectories confined to the central tile, gain parameter directly modulated

The central tile of the firing rate phase plane, where both functions $f^{pw}_{g(t)}(\cdot)$ appearing in Eqs. (15)–(16) are linearly increasing, is defined by $\beta y_1 \in [a_2(t)$

$-b-(1/2g(t)), a_2(t)-b+(1/2g(t))]$ and $\beta y_2 \in [a_1(t)-b-(1/2g(t)), a_1(t)-b+(1/2g(t))]$. If

$$|a_1(t)-b| < \frac{1}{2g(t)}, \quad |a_2(t)-b| < \frac{1}{2g(t)}, \qquad (20)$$

then the central tile always contains the origin and some part of the first quadrant (note that this quadrant is invariant under the deterministic part of Eqs. (7)–(8) if f is non-negative) so that decision dynamics starting at the origin may (for suitable choices of other parameters) take place entirely within the central tile. For example, if $b = 0.5$ and $0 < g(t) \leq 1$, then $a_1(t)$, $a_2(t)$ may take values between 0 and 1 while still satisfying (20).

Figure 4 shows a sample of solutions of the piecewise-linearized firing rate model for the piecewise constant parameters $g(t) = \{0.3,\ t < t_s; 1,\ t \geq t_s\}$, $a_1(t) = \{1, t \leq t_s; 1.03, t > t_s\}$, $a_2(t) = \{1, t \leq t_s; 0.97, t > t_s\}$, $c(t) \equiv 0.09\sqrt{2}$, $b = 0.5$, $\tau_c = 1$, $\theta = 0.725$, $t_s = 10$ and $\beta = 1$. Note that stimuli $a_j(t) \neq 0$ are present throughout, but that coherence ($a_1(t) \neq a_2(t)$) appears in the inputs a_j only at $t = t_s$, so that times $t < t_s$ make up the preparatory phase mentioned in the introduction and the situation corresponds to the introduction of coherence into an entirely random pattern. Assuming that decision thresholds are set within the boundaries of the central tile or that the interrogation time T is sufficiently small so that only a negligible proportion of solutions have left this tile, solutions are effectively confined to the central tile for all times of interest. This behavior characterizes Case 1 parameter sets, for which subtraction of Eqs. (15)–(16) yields the one-dimensional SDE

$$\tau_c \frac{dy}{dt} = -y + g(t)(\beta y + a(t)) + g(t)c(t)\eta_t \qquad \text{(firing rate model)}, \qquad (21)$$

where we define the net rate of incoming evidence as

$$a(t) = a_1(t) - a_2(t). \qquad (22)$$

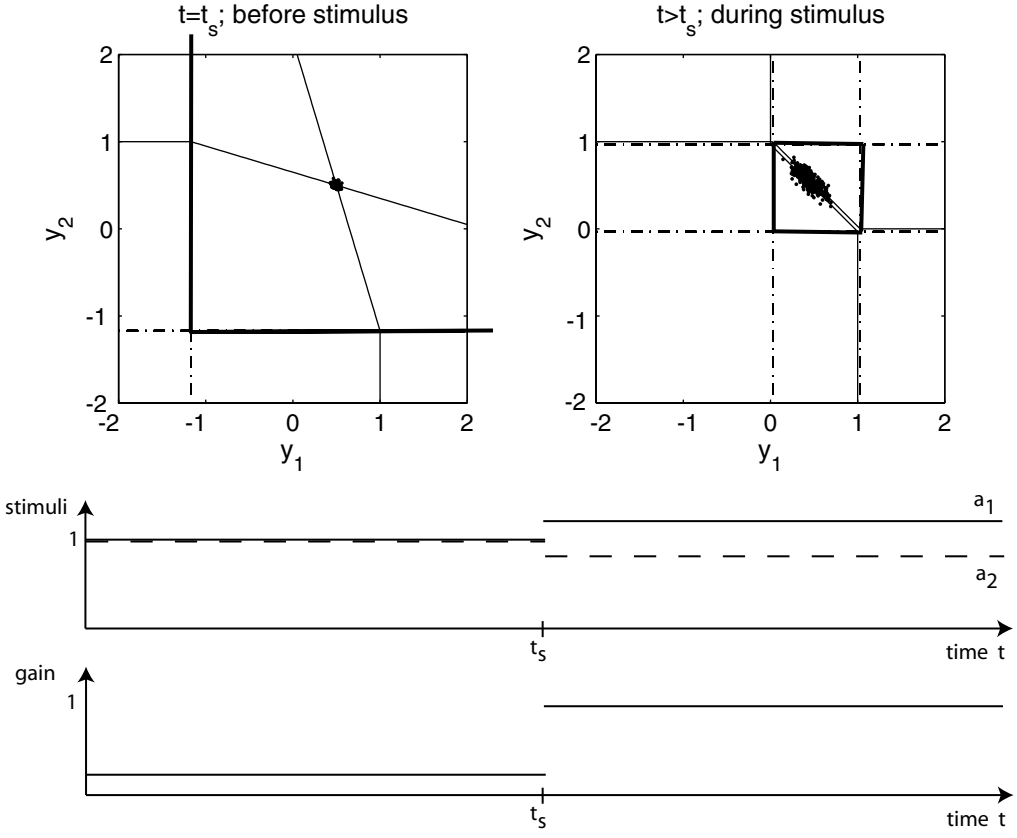


Fig. 4. Case 1: solutions confined to central tile. Scatter plot of trajectories both at the end of the preparatory period and hence at the moment of stimulus onset t_s (left) and during the stimulus ($t = t_s + 2$, right). The tiling of the plane is shown with dot-dashed lines; cf. Fig. 2; the central tile is outlined in solid and extends outside the plotted domain in the left panel. Parameter values are given in text. Also shown are nullclines for Eqs. (15)–(16) as thin solid lines. The lower panels show stimuli $a_j(t)$ and gain $g(t)$ as functions of time.

We note that transient gain values in this case result from modifications to the firing rate function itself, as solutions explore only the central region of this function in which it is practically linear. This is the "external" mechanism of dynamic gain change discussed in the Introduction.

For future reference, we also note that an analytical expression for the density of reaction times may be derived if the parameters in (21) are constant (i.e. $a(t) \equiv a$, $c(t) \equiv c$) and the gain "balances" the decay: e.g. $g(t) \equiv g = 1$ in (21) (see, e.g. [Ratcliff *et al.*, 1999]). In this case, (21) simplifies to a constant drift diffusion process and the probability that a trajectory first escapes the interval $[-\overline{\theta}, \overline{\theta}]$ at a time $RT = \inf\{t : |y(t)| > \overline{\theta}\}$ from initial condition $y(0) = 0$ has density

$$p(RT) = \frac{\pi c^2}{\overline{\theta}^2} e^{-\frac{a^2 RT}{2c^2}} \left(e^{-\frac{\overline{\theta} a}{c^2}} + e^{\frac{\overline{\theta} a}{c^2}} \right) \times \sum_{k=1}^{\infty} k \sin\left(\frac{k\pi}{2}\right) \exp\left(\frac{-k^2\pi^2 c^2 RT}{8\overline{\theta}^2}\right). \tag{23}$$

Here $\pm\overline{\theta}$ correspond to the intersections of the decision manifold $\mathcal{M}$ with the thresholds $y_j = \theta$ of the two-dimensional process (Fig. 3). Equation (23) may be extended to account for distributed initial conditions $y(0) \neq 0$ and other generalizations [Ratcliff *et al.*, 1999], but we do not use such extensions here.

Similar considerations yield the reduction of the connectionist model restricted to its respective central tile:

$$\tau_c \frac{dx}{dt} = -x + \beta g(t) x + a(t) + c(t)\eta_t \quad \text{(connectionist model)}. \tag{24}$$

Note that gain multiplies the last three terms in (21), but only the second in (24).

Case 2. Trajectories switch tiles, changing effective gain

We now consider the case of stimuli $a_j(t)$ that "suddenly" turns on from zero at time t_s while the gain parameter $g(t) \equiv g$ remains constant, and show how stimulus onset itself can give rise to a time-dependent one-dimensional reduction that resembles the reduction to (21) obtained above. This corresponds to appearance of a partially coherent stimulus replacing a fixation spot. Since $a_1(t) = a_2(t) = 0$ for $t \leq t_s$, in this period there is a stable fixed point at $(0, 0)$ if $b \geq 1/2g$. If $b = 1/2g$, the situation simplifies: while $t \leq t_s$, $(0, 0)$ lies exactly at the corner of tile 9 (see Fig. 2), to which tile solutions are confined (modulo noise effects). At stimulus onset t_s, tile boundaries shift, so that, for appropriate choices of $a_1(t), a_2(t) > 1/2g(t) - b$ for $t > t_s$, the origin and the cluster of solutions in its neighborhood at time $t = t_s^+$, suddenly finds itself in the central tile 5. For concreteness, we fix parameters meeting the requirements $b = 1/2g$ and $a_1(t) = a_2(t) = 0$ for $t \leq t_s$ as follows: $a_1(t) = \{0, t \leq t_s; 1.03, t > t_s\}$, $a_2(t) = \{0, t \leq t_s; 0.97, t > t_s\}$, $g = 1$ and all other parameters as for the example in Case 1. See Fig. 5.

To determine the appropriate linear (two- and one-dimensional) reductions for these parameters, we use Eqs. (15)–(16) restricted to tile 9 for the preparatory phase $t \leq t_s$, and restricted to tile 5 for times $t > t_s$ during stimulus presentation (we make the same assumptions about the interrogation time or thresholds as for Case 1, so that solutions remain in the central tile 5 for all times $t > t_s$ of relevance to the decision). This yields the one-dimensional equation

$$\tau_c \frac{dy}{dt} = -y + \begin{cases} gc(t)\eta_t & \text{for } t \leq t_s \\ g[\beta y + a(t)] + gc(t)\eta_t & \text{for } t > t_s \end{cases}, \tag{25}$$

(and an analogous reduction to a linear two-dimensional model).

Equation (25) is similar to the reduction (21), if the stimulus and gain functions in the latter are piecewise constant, as for the example parameters of Case 1. The major difference is that the noise coefficient remains constant for (25). As we see in the next section, the statistics produced by the one-dimensional models (21) and (25) can nevertheless agree rather well. Thus, transient gain strategies to be derived for the more general (21) in Sec. 3 can be approximately implemented for stimuli undergoing large steps, with no changes in the gain of the activation functions *per se*.

Similar considerations hold for Cases 1 and 2 reductions of the connectionist model, but we do not pursue this here.

2.6. *Accuracy of the reduced models*

Figure 6 demonstrates that our simplifications of the nonlinear firing rate model (7)–(8) accurately capture reaction time statistics for Case 1

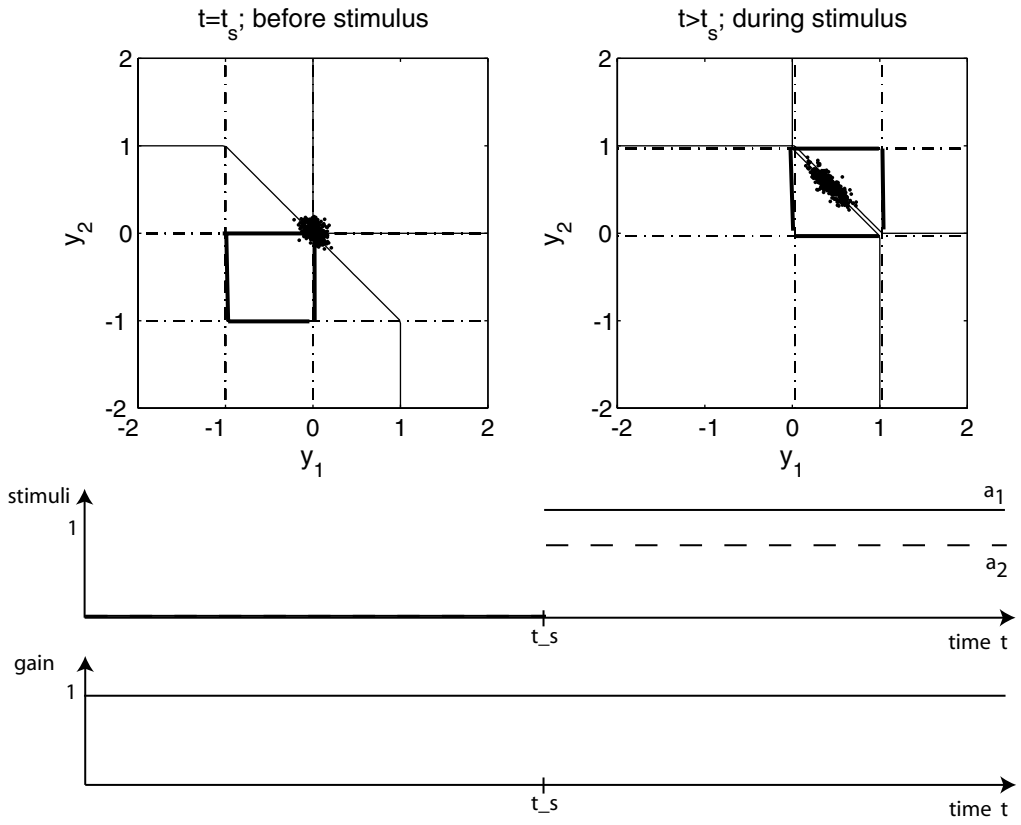
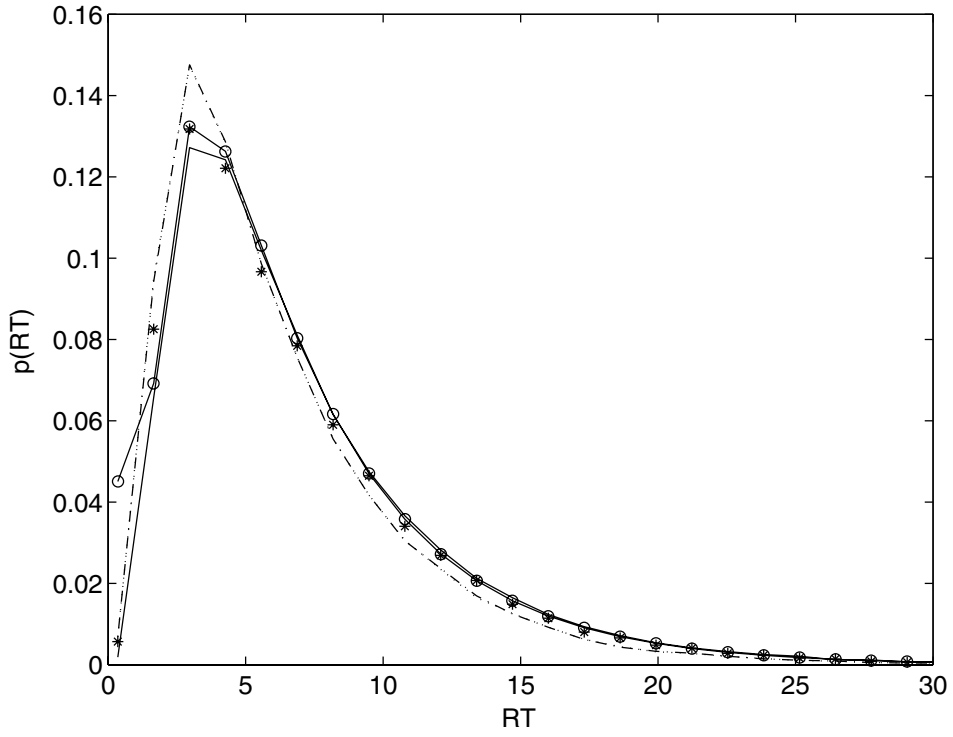


Fig. 5. Case 2: trajectories switch tiles. Scatter plot of trajectories both at the end of the preparatory period and hence at the moment of stimulus onset t_s (left) and during the stimulus ($t = t_s + 2$, right). The tiling of the plane is shown with dot-dashed lines; cf. Fig. 2; the central tile is outlined in solid. Parameter values are given in text. Also shown are nullclines for Eqs. (15)–(16) as thin solid lines. The lower panels show stimuli $a_j(t)$ and gain $g(t)$ as functions of time.

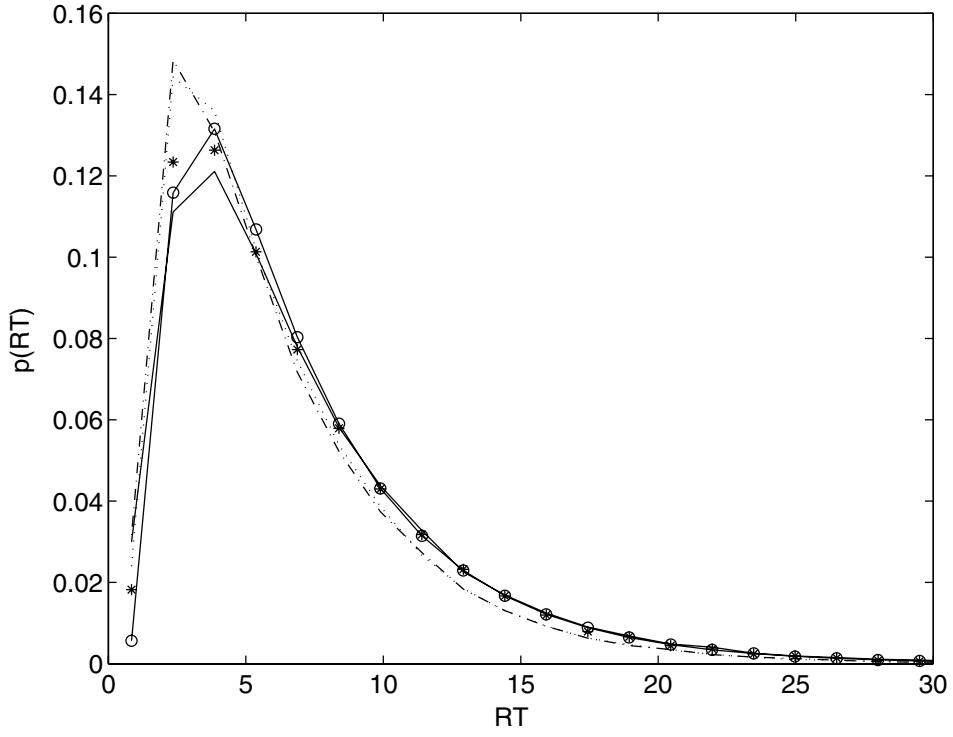
parameters. For the one- and two-dimensional linear reductions, linearized activation functions take piecewise constant (in time) values appropriate to the tiles containing the dominant proportion of solution trajectories during the preparatory and trial periods, exactly as in (21). That is: for Case 1, $f^l_{g(t)}(x) = 1/2 + (x - b)$ for all t, as solutions remain in the central tile 5. For Case 2, $f^l_{g(t)}(x) = 0$ for all $t < t_s$ (when solutions are in tile 9) and $f^l_{g(t)}(x) = (1/2) + (x - b)$ for $t > t_s$, when solutions are in tile 5.

For Case 1, the error rates corresponding to the reaction time distributions of Fig. 6 are 0.050, 0.051, 0.051, 0.035, and 0.034 respectively for the two-dimensional firing rate model with logistic activation functions $f_{g(t)}$, the two-dimensional model with piecewise-linear activation functions $f^{pw}_{g(t)}$ (15)–(16), the two-dimensional model with linear activation functions, the one-dimensional reduction (21), and the expression (23), which describes the one-dimensional reduction with initial condition $y(t_s) = 0$ at the time of stimulus presentation (keeping the first 10 terms of sum). For Case 2, these error rates are 0.060, 0.065, 0.059, 0.042, 0.034. Thus, in both cases while the different two-dimensional models are in close agreement, the one-dimensional reductions produce significantly lower error rates. Figures 4 and 5 show why: the distribution of solutions is not entirely collapsed along the attracting decision manifold, and the spatially extended "incorrect" thresholds of the two-dimensional models require smaller (and hence more probable) excursions to cross. Closer agreement between one- and two-dimensional models can be achieved with, for example, higher values of β or lower values of noise strength c: see [Bogacz *et al.*, 2004].

As an additional comparison among the various models, we separately computed error rates for interrogation at a time $T = t_s + 1$ (see [Usher & McClelland, 2001] for an earlier, related comparison between the nonlinear two-dimensional and linear one-dimensional models). For Case 1,



(a)



(b)

Fig. 6. Reaction time densities for the nonlinear firing rate model of Eqs. (7)–(8) (stars) and its various reductions, with thresholds $\theta = 0.725$: dot-dashed line, two-dimensional model with piecewise-linear activation functions $f^{pw}_{g(t)}$; dotted line, two-dimensional model with linear activation functions $f^{l}_{g(t)}$; solid line, linear one-dimensional reduction, solid line with circles, analytic 1-D expression with zero variance at trial onset (see text). (a) Case 1: solutions confined to central tile. (b) Case 2: trajectories switch tiles. Parameter values are given in main text.

the interrogation error rates are (in the same order as above) 0.323, 0.321, 0.321, 0.324, and 0.319. For Case 2, these error rates are 0.374, 0.363, 0.354, 0.350, and 0.319. For both cases interrogation error rates are more similar for the various model reductions than the free response error rates reported in the previous paragraph. This is expected from the discussion in Sec. 2.5, since accurate description of the interrogation protocol by a one-dimensional model does not require that solutions are confined near the decision manifold.

2.7. *Drift-diffusion and the one-dimensional models as linear filters*

We introduce a third one-dimensional SDE, an extension of the drift-diffusion model of [Laming, 1968; Ratcliff, 1978] in which both drift and diffusion terms are multiplied by a common gain factor $g(t)$:

$$\tau_c \frac{dz}{dt} = g(t)[a(t) + c(t)\eta_t] \qquad \text{((pure) drift-diffusion model).} \tag{26}$$

Equation (26) and the one-dimensional reductions of the firing rate and connectionist equations (21) and (24) are Ornstein–Uhlenbeck processes, (affine-) linear in the activities x, y and z and in the input

$$\underbrace{I(t)}_{\text{input}} = \underbrace{a(t)}_{\text{signal}} + \underbrace{c(t)\eta_t}_{\text{noise}} . \tag{27}$$

We may explicitly solve all these SDEs, for a given realization of the white noise process η_s, $s \in [0, t]$, to obtain respectively

$$z(t) = \int_0^t \frac{g(s)a(s)}{\tau_c} ds + \int_0^t \frac{g(s)c(s)}{\tau_c} dW_s \tag{28}$$

for the drift diffusion model,

$$x(t) = \int_0^t \frac{a(s)}{\tau_c} \exp\left(\frac{1}{\tau_c}\int_s^t [\beta g(s') - 1]\, ds'\right) ds + \int_0^t \frac{c(s)}{\tau_c} \exp\left(\frac{1}{\tau_c}\int_s^t [\beta g(s') - 1]\, ds'\right) dW_s \tag{29}$$

for the connectionist model, and

$$y(t) = \int_0^t \frac{a(s)g(s)}{\tau_c} \exp\left(\frac{1}{\tau_c}\int_s^t [\beta g(s') - 1]\, ds'\right) ds + \int_0^t \frac{c(s)g(s)}{\tau_c} \exp\left(\frac{1}{\tau_c}\int_s^t [\beta g(s') - 1]\, ds'\right) dW_s \tag{30}$$

for the firing rate model. Here, dW_s is an increment of a Wiener process, of which the white noise process η_s is the formal time derivative, and we have assumed unbiased initial data $x(0) = y(0) = z(0) = 0$. These expressions all take the form

$$w(t) = \int_0^t K(t, s)a(s)ds + \int_0^t K(t, s)c(s)dW_s\,, \tag{31}$$

and so we conclude that (28)–(30) all compute linear filters of their inputs.

At any fixed time t, $w(t)$ is a Gaussian-distributed random variable with mean $\int_0^t K(t, s)a(s)$ and variance $\int_0^t K^2(t, s)c^2(s)ds$. Using this fact, after a change of variables the error rate expression (19) becomes

$$\mathrm{ER} = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{\left|\int_0^t K(t, s)a(s)\right|}{\sqrt{\int_0^t K^2(t, s)c^2(s)ds}}\right)\right] . \tag{32}$$

3. Optimal Signal Discrimination in the One-Dimensional Models

We now ask what functional form of $g(t)$ optimizes performance for Eqs. (28)–(30), thereby computing optimal gain trajectories for the (reduced) drift-diffusion, connectionist, and firing rate models.

3.1. *Optimal statistical tests*

Given only the noisy input function (27), consider the task of deciding whether $I(t)$ was generated by time-dependent signals $a_0(t)$ or $a_1(t)$: hypotheses 0 and 1, resp. This can be accomplished in two distinct ways, mirroring the interrogation and free response protocols of Sec. 2. In the first, the decision is made at a fixed time T; in the second, it is made when some preset level of confidence is reached. Optimal performance in the first version of the task implies that as few errors as possible are made; in the second, it implies that

the decision must be made as quickly as possible for a fixed error tolerance, timed from stimulus onset at time $t = 0$. The best strategy in the first version is the (continuum limit of the) Neyman–Pearson test; in the second version it is the sequential probability ratio test (SPRT) [Wald, 1947; Lehmann, 1959]. Both tests compute an evolving estimate of the log likelihood ratio:

$$l(t) = \log\left[\frac{p(\{I(s)|a_0(s), s \in [0,t]\})}{p(\{I(s)|a_1(s), s \in [0,t]\})}\right] \triangleq \log\left[\frac{p_0(\{I(s), s \in [0,t]\})}{p_1(\{I(s), s \in [0,t]\})}\right]. \quad (33)$$

(the base of the logarithm is arbitrary). In the Neyman–Pearson test, hypothesis 0 is chosen if $l(T) > 0$ and hypothesis 1 if $l(T) < 0$; in the SPRT, hypothesis 0 (resp. 1) is chosen when $l(t)$ first crosses threshold θ (resp. $-\theta$), θ being determined by the error tolerance.

Writing the input $I(t)$ (27) as a sum of its increments for an appropriate discretization of time $\{t^j\}$:

$$I(t) = \sum_j dI^j = \sum_j a(t^j)dt + c(t^j)\, dW_t^j, \quad (34)$$

we obtain

$$l(t) = \sum_j \log\left[\frac{p_0(dI^j)}{p_1(dI^j)}\right]. \quad (35)$$

Now restrict to the special case in which $a_0(t) = -a_1(t) = a(t)$ and consider the likelihood distributions (now themselves time-dependent) that correspond to an increment $dI(t) = a(t)dt + c(t)dW_t$. Since the dW_t are normally distributed with mean 0 and variance dt, we have

$$p_0(t)(dI(t)) = \frac{1}{\sqrt{2\pi c^2(t)dt}}\, e^{-(dI(t)+a(t)dt))^2/(2c^2(t)dt)}, \quad (36)$$

$$p_1(t)(dI(t)) = \frac{1}{\sqrt{2\pi c^2(t)dt}}\, e^{-(dI(t)-a(t)dt))^2/(2c^2(t)dt)}. \quad (37)$$

The corresponding increment of likelihood evidence to (33) is

$$dl_t = \log\left(\frac{p_1(dI_t)}{p_0(dI_t)}\right) = k\frac{a(t)}{c^2(t)}dI_t\,, \quad (38)$$

where $k = 2\log(e)$ depends on the base of the logarithm. Substituting for dI_t, we obtain a differential equation for the total evidence l_t accumulated at time t,

$$dl_t = k\left[\frac{a^2(t)}{c^2(t)}\,dt + \frac{a(t)}{c(t)}\,dW_t\right], \quad (39)$$

which may be integrated to yield:

$$l(t) = \int_0^t k\frac{a^2(s)}{c^2(s)}ds + \int_0^t k\frac{a(s)}{c(s)}dW_s. \quad (40)$$

Comparing with Eq. (31) shows that the optimal filter is

$$K(t,s) = k\frac{a(s)}{c^2(s)}: \quad (41)$$

this is the matched filter for white noise which is fundamental in signal processing [Papoulis, 1977]. Note that, in (39)–(40) only the signal-to-noise ratio (a/c) appears.

3.2. *A direct proof that the kernel $K(t,s) = k(a(s)/c^2(s))$ is optimal in the interrogation paradigm*

As follows from its matched filter property, the linear filter $K(t,s) = k(a(s)/c^2(s))$ which computes log likelihood $l(t)$ for inputs with white noise also produces, for all times t, a filtered (and Gaussian) version $w(t)$ of the input [Eq. (31)] with a maximal integrated signal-to-noise ratio

$$F[K;a,c](t) = \frac{\left|\int_0^t K(t,s)a(s)ds\right|}{\sqrt{\mathbb{E}\left(\int_0^t K(t,s)c(s)dW_s\right)^2}} = \frac{\left|\int_0^t K(t,s)a(s)ds\right|}{\sqrt{\int_0^t K^2(t,s)c^2(s)ds}}. \quad (42)$$

For completeness, we now demonstrate this directly.

Minimization of the error rate (18) or (19) for (fixed) interrogation at time $t = T$ is achieved by maximizing F over all possible kernels $K(s)$. This problem in the calculus of variations is solved by computing the first and second variations, with respect to K, of the functional F, setting the first to zero to determine a candidate $\overline{K}$ for the optimal K, and evaluating the second at $\overline{K}$ to check that $D_K^2 F$ is negative (semi-) definite. Henceforth we drop explicit reference to the (fixed, arbitrary) interrogation time $t = T$ in the function K and

write $K(T,s) = K(s)$. We compute:

$$\frac{\delta F}{\delta K} = \lim_{\epsilon\to 0}\frac{d}{d\epsilon}F[K+\epsilon\gamma;a,c](T) = \lim_{\epsilon\to 0}\frac{d}{d\epsilon}\left\{\frac{\displaystyle\int_0^T a(s)[K(s)+\epsilon\gamma(s)]\,ds}{\left[2\displaystyle\int_0^T c^2(s)[K^2(s)+2\epsilon g(s)\gamma(s)+\epsilon^2\gamma^2(s)]\,ds\right]^{\frac{1}{2}}}\right\}$$

$$= \lim_{\epsilon\to 0}\frac{1}{\sqrt{2}}\left\{\frac{\displaystyle\int_0^T a(s)\gamma(s)\,ds}{[H(T,\epsilon)]^{\frac{1}{2}}} - \frac{\displaystyle\int_0^T a(s)[K(s)+\epsilon\gamma(s)]\,ds\int_0^T c^2(s)[K(s)\gamma(s)+\epsilon\gamma^2(s)]\,ds}{[H(T,\epsilon)]^{\frac{3}{2}}}\right\}$$

$$= \frac{\displaystyle\int_0^T a(s)\gamma(s)\,ds\int_0^T c^2(s)K^2(s)\,ds - \int_0^T a(s)K(s)\,ds\int_0^T c^2(s)K(s)\gamma(s)\,ds}{\sqrt{2}\left[\displaystyle\int_0^T c^2(s)K^2(s)\,ds\right]^{\frac{3}{2}}}, \tag{43}$$

where $H(T,\epsilon) = \int_0^T c^2(s)[K^2(s) + 2\epsilon K(s)\gamma(s) + \epsilon^2\gamma^2(s)]\,ds$. Setting (43) equal to zero and using the fact that the variation $\gamma(s)$ is arbitrary, we conclude that the critical point indeed occurs at $\overline{K}(s) = k(a(s)/c^2(s))$, as given by (41).

To compute the second derivative we differentiate the expression within braces in the penultimate step of (43) with respect to ϵ once more, set $\epsilon = 0$, and evaluate the resulting expression at the critical point (41), obtaining:

$$\left.\frac{\delta^2 F}{\delta K^2}\right|_{K=\overline{K}} = -\frac{\displaystyle\int_0^T c^2(s)\overline{K}^2(s)\,ds\int_0^T c^2(s)\gamma^2(s)\,ds - \left(\int_0^T c^2(s)\overline{K}(s)\gamma(s)\,ds\right)^2}{\sqrt{2}\left[\displaystyle\int_0^T c^2(s)\overline{K}^2(s)\,ds\right]^{\frac{3}{2}}} \le 0\,. \tag{44}$$

In the last step we appeal to Schwarz's inequality. This proves that the second variation is negative semidefinite, and vanishes identically only for variations $\gamma(s) = \kappa\overline{K}(s)$ in the direction of $\overline{K}$ (as expected from (41), which contains the arbitrary "scaling" parameter k).

Substituting (41) into (42) we obtain

$$F[\overline{g};a,c](T) = \sqrt{\frac{1}{2}\int_0^T \frac{a^2(s)}{c^2(s)}\,ds}\,, \tag{45}$$

and using (32), we obtain the minimum possible error rate for interrogation at time t:

$$\mathrm{ER} = \frac{1}{2}\left[1 - \mathrm{erf}\left(\sqrt{\frac{1}{2}\int_0^T \frac{a^2(s)}{c^2(s)}\,ds}\right)\right]. \tag{46}$$

Since the integrand $(a/c)^2$ is non-negative, the error rate continues to decrease or at worst remains constant as T increases.

3.3. *Optimal gains for the three models*

We may now extract explicit expressions for optimal gains by setting $K(s) = \overline{K}(s)$ in (31) and comparing the resulting integrands with those in the SDE solutions (28)–(30).

3.3.1. *Pure drift-diffusion model*

Comparing (31) with (28), we see that the optimal gain is simply $\overline{K}$:

$$\overline{g}_{dd}(s) = \tau_c\overline{K}(s) = \tau_c k\frac{a(s)}{c^2(s)}\,; \tag{47}$$

thus, there is a continuum of optimal schedules differing only by a multiplicative scale factor.

3.3.2. *Connectionist model*

Equations (31) and (29) give

$$\tau_c \overline{K}(s) = \tau_c k \frac{a(s)}{c^2(s)} = \exp\left(\frac{1}{\tau_c}\int_s^T [\beta \overline{g}_c(s') - 1]\, ds'\right), \quad (48)$$

where $\overline{g}_c$ is the optimal gain for the connectionist model. Taking the log of this expression, differentiating with respect to s, and solving for $\overline{g}_c(s)$, we obtain:

$$\overline{g}_c(s) = \frac{1}{\beta}\left[1 - \tau_c \frac{d}{ds}\log\left(\frac{a(s)}{c^2(s)}\right)\right]. \quad (49)$$

Note that $\overline{g}_c$ is unique and in particular, independent of k and the interrogation time T. However, $\overline{g}_c$ is not required to be positive, so may not always be physically admissable. The form of $\overline{g}_c$ may be interpreted as follows. When $(a(s)/c^2(s))$ is decreasing, $\overline{g}_c(s) > 1/\beta$ and the O–U process (24) is unstable; hence solutions "run away," in the direction $x(s)$, emphasizing higher-fidelity information that was previously collected. When $(a(s)/c^2(s))$ is increasing, $\overline{g}_c(s) < 1/\beta$, the O–U process is stable, and the linear term in (24) is attractive, thereby discounting previously integrated information in favor of the higher-fidelity input currently arriving.

We note that, because the "output" neural activity is determined by a gain-dependent function of the dynamical variable x in the connectionist model (see text following Eqs. (1)–(2)), transient gain schedules also adjust the position of free-response thresholds with respect to x. We leave an exploration of this effect, which does not enter the interrogation protocol or affect the firing rate model, for future studies.

3.3.3. *Firing rate model*

Equations (31) and (30) give

$$\tau_c \overline{K}(s) = \tau_c k \frac{a(s)}{c^2(s)} = \overline{g}_f(s)\exp\left(\frac{1}{\tau_c}\int_s^T [\beta \overline{g}_f(s') - 1]\, ds'\right). \quad (50)$$

Defining $f(s) = \tau_c k(a(s)/c^2(s))e^{(1/\tau_c)(T-s)}$, differentiating with respect to s, and restricting to positive functions $\overline{g}_f$, a and c^2 (which we justify below), (50) yields

$$\begin{aligned} f'(s) &= \frac{d}{ds}\left[\overline{g}_f(s)\exp\left(\frac{1}{\tau_c}\int_s^T \beta \overline{g}_f(s')ds'\right)\right] \\ &= \overline{g}'_f(s)\exp\left(\frac{1}{\tau_c}\int_s^T \beta \overline{g}_f(s')ds'\right) \\ &\quad - \frac{\beta}{\tau_c}\overline{g}_f^2(s)\exp\left(\frac{1}{\tau_c}\int_s^T \beta \overline{g}_f(s')ds'\right) \\ &= \overline{g}'_f(s)\frac{f(s)}{\overline{g}_f(s)} - \frac{\beta}{\tau_c}\overline{g}_f(s)f(s). \end{aligned} \quad (51)$$

Rewriting (51), we obtain

$$\begin{aligned} \frac{d\overline{g}_f(s)}{ds} &= \frac{\beta}{\tau_c}\overline{g}_f^2(s) + \overline{g}_f(s)\frac{f'(s)}{f(s)} \\ &= \frac{\beta}{\tau_c}\overline{g}_f^2(s) + \overline{g}_f(s)\frac{d}{ds}\log(f(s)) \\ &= \frac{\beta}{\tau_c}\overline{g}_f^2(s) + \overline{g}_f(s)\left[\frac{d}{ds}\log\left(\frac{a(s)}{c^2(s)}\right) - \frac{1}{\tau_c}\right]. \end{aligned} \quad (52)$$

Thus, the condition for optimal gain in the linearized firing rate model is a differential equation, unlike the algebraic relationships for the drift-diffusion and connectionist cases. Note that solutions to (52) initialized at positive values remain positive for all times, since the equation has an equilibrium at $\overline{g}_f = 0$, preventing passage through this point. This justifies our assumption of positive $\overline{g}_f$ above and ensures that the optimum gain is "physical" this sense. In fact, (52) may be solved explicitly using the integrating factor $I(s) = \exp\left(\int_0^s l(s')ds'\right)$, where $l(s') \triangleq (d/ds')\log(a(s')/c^2(s')) - 1/\tau_c$, yielding

$$\overline{g}_f(s) = \frac{\exp\left(\int_0^s l(s')ds'\right)}{\frac{\beta}{\tau_c}\int_0^s \left[\exp\left(\int_0^{s'} l(s'')ds''\right)\right]ds' + \frac{1}{g(0)}}. \quad (53)$$

The integral equation (50) specifies only an *arbitrary*, positive final condition $\overline{g}_f(T) = k(a(T)/c^2(T))$ for (52), since k is itself arbitrary. Any solution of (52) with positive initial condition

(as long as it is defined) therefore delivers a member of the continuum of optimal gain functions for the linearized firing rate model. This is in striking contrast to the unique optimal gain (49) in the connectionist model, and, since the different $\overline{g}_f$ generally have different forms (see below), it also contrasts with the multiplicity of "scaled" optimal drift-diffusion gain functions (47). The optimality of $\overline{g}_f$ schedules with such different forms follows from the fact that gain multiplies the inputs to the firing rate model (21). For example, optimal gain schedules with $(\beta\overline{g}_f(s) - 1) < 0$ may implement the SPRT even when the signal-to-noise-ratio is constant (see Example 1 below), because discounting of previously integrated evidence is compensated for via weighting incoming evidence by a decreasing function $\overline{g}_f(s)$.

3.3.4. *Numerical examples*

Example 1. We first take constant signal $a(s) \equiv a = 0.06$ and constant noise strength $c(s) \equiv 0.09$ with $\tau_c = \beta = 1$. Then, Eq. (47) gives the family of optimal constant gain functions for the pure drift-diffusion model,

$$\overline{g}_{dd}(s) \equiv \tau_c k a\,, \tag{54}$$

and Eq. (49) gives the unique optimal gain for the connectionist model, again a constant:

$$\overline{g}_c(s) \equiv \frac{1}{\beta}\,. \tag{55}$$

For the same parameter values, the firing rate model gain ODE (52) becomes

$$\frac{d}{ds}\overline{g}_f(s) = \frac{\beta}{\tau_c}\overline{g}_f^2(s) - \frac{1}{\tau_c}\overline{g}_f(s). \tag{56}$$

Initial conditions $\overline{g}_f(0) \in [0, 1/\beta]$ decay to the fixed point at $\overline{g}_f = 0$, while for $\overline{g}_f(0) > 1/\beta$, gain functions increase to ∞ in finite time. The initial condition $\overline{g}_f(0) = 1/\beta$ yields the constant gain function $\overline{g}_f(s) \equiv 1/\beta$, for which the linearized firing rate model again becomes constant drift Brownian motion: see Fig. 7. As expected, all gain profiles

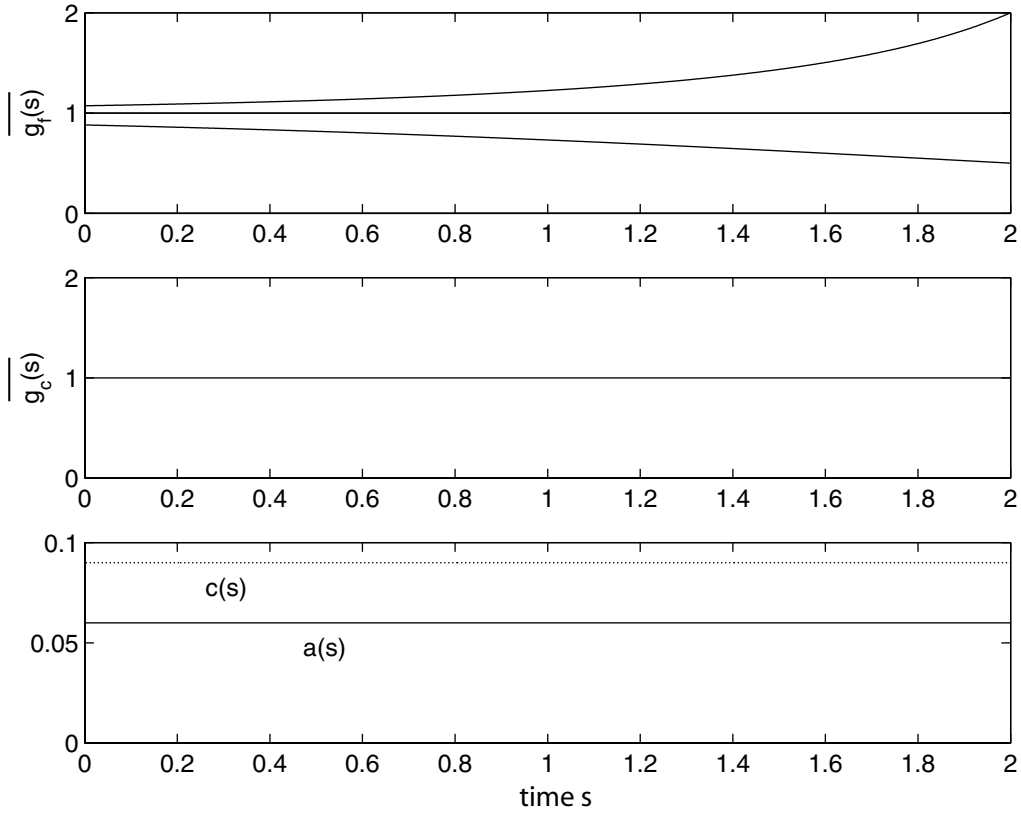


Fig. 7. Optimal gains for constant signal strength $a(s) \equiv 0.06$ (solid line in bottom panel) and constant noise amplitude $c(s) \equiv 0.09$ (dotted line). Top panel: three optimal gain schedules $\overline{g}_f$ solving (52); note that these include, but are not limited to, $\overline{g}_f(s) \equiv 1/\beta$ (here $\beta = 1$). Central panel: the unique optimal gain function $\overline{g}_c(s) \equiv 1/\beta$ for the connectionist model, given by Eq. (49).

produced optimal performance (with 82.7% correct responses returned at interrogation time $T = 2$).

Example 2. We now assume that signal amplitude is zero up to stimulus presentation at time t_s and rises exponentially toward $\overline{a}$ thereafter: $a(s) = \overline{a}[1 - e^{-r(s-t_s)}]$ for $s > t_s$. This form is motivated by the saturating dynamics of input layers which feed forward to decision units in simple connectionist models. We set $\overline{a} = 0.06$, $r = 10$, $t_s = 1$ and take constant noise strength $c(s) \equiv 0.09$ and $\tau_c = \beta = 1$ as previously: see Fig. 8 (bottom). As $r \to \infty$, $a(s)$ approaches the piecewise constant functions of Secs. 2.5.1–2.6, for which the one-dimensional reduction was shown to be an adequate model.

For the pure drift-diffusion model, Eq. (47) gives

$$\overline{g}_{dd}(s) = \tau_c k a(s), \tag{57}$$

so that, as above, optimal gain trajectories are scaled versions of the signal strength and, in particular, $\overline{g}(s) = 0$ for $s \le t_s$. For the connectionist and firing rate models, however, the formulae (49) and (52) are valid only while $a(s) > 0$, and additional reasoning is needed to determine optimal gain values in the pre-stimulus period $s < t_s$. For the connectionist model, the integral equation (48) is clearly satisfied for $a(s) = 0$ if $g_c(s) = -\infty$, so we set $\overline{g}_c(s) = -\infty$, $s \le t_s$. Since for a "physical" neural network, activation functions $f_{g(t)}(\cdot)$ are nondecreasing, such negative gain values are not directly relevant to biological applications, but illustrate the demand that relative activation x be clamped at zero before the stimulus arrives. As before, we define $\overline{g}_c(s)$ via (49) for $s > t_s$. That is, for $t > t_s$,

$$\overline{g}_c(s) = \frac{1}{\beta}[1 - \tau_c l(s)]\,, \tag{58}$$

where $l(s) = (d/ds)\log(a(s)/c^2(s)) = r/e^{r(s-t_s)} - 1$ decays from ∞ to 0 as time s increases.

For the firing rate model, we also appeal directly to the integral equation (50) to define $g_f(s)$ when $a(s) = 0$. Since (50) is satisfied by $\overline{g}_f(s) = 0$,

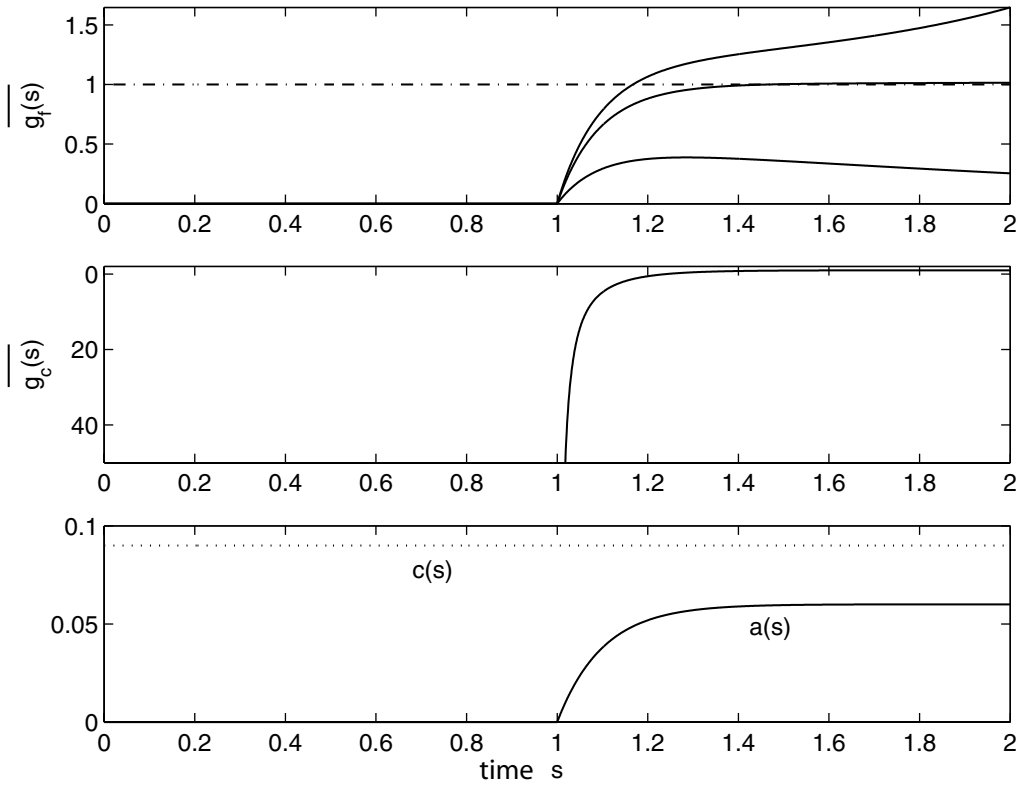


Fig. 8. Optimal gains for exponentially asymptoting signal strength $a(s)$ (solid line in bottom panel) and constant noise amplitude $c(s) \equiv 0.09$ (dotted line). Top panel: three optimal gain schedules $\overline{g}_f$ for the firing rate model solving (52) (solid curves); the nonoptimal constant gain $g \equiv 1/\beta$ is shown as dot-dashed for reference. The lowest of the solid $\overline{g}_f$'s displays the rise-decay form discussed in the text. Central panel: the unique optimal gain function for the connectionist model, given by Eq. (49); $\overline{g}_c(s) = -\infty$ for $s \le t_s$.

we assume this for $s \le t_s$. We then determine $\overline{g}_f(s)$ for $s > t_s$ from (52), allowing a discontinuity at t_s and taking arbitrary "initial" conditions $\overline{g}_f(t_s)$. Figure 8 illustrates several optimal functions arising from different choices of $\overline{g}_f(t_s)$. The following fact is helpful in understanding positive solutions of (52): orbits lying below $(1/\beta)[1 - \tau_c\, l(s)]$ at any time s decrease toward 0; those above this value increase. Since $(1/\beta)[1 - \tau_c\, l(s)] \to 1/\beta$ as $s \to \infty$, $1/\beta$ asymptotically forms a separatrix between optimal gain trajectories that decay and those that diverge to ∞. Also, note that Case 2 parameters for the two-dimensional firing rate model of Sec. 2.5.1 implement a step in effective gain values up to $1/\beta = 1$, so that in this case nearly optimal signal processing occurs with no explicit adjustment of the gain parameter. The performance resulting from optimal gain trajectories in all models is 73.1% correct responses at interrogation at time $T = 2$; for comparison, the (*non*optimal) constant gain $\overline{g}_f(s) \equiv 1/\beta$ produces only 66.4% correct.

Gains must remain bounded for all time to be of practical interest. A family of optimal gain schedules of this form, determined by their (sufficiently small) initial conditions, will always exist for monotonically rising and bounded stimuli $a(s)$ such as that chosen here. As we elaborate in Sec. 4, their "rise-decay" pattern resembles the gain produced by dissipating pulses of the neuromodulator norepinephrine delivered to cortical decision areas via the locus coeruleus, hence providing a clue that this brainstem organ may be assisting near-optimal decision making.

Example 3. We finally assume that $a(s)$ smoothly increases from a low to a higher level and then returns to its original level, corresponding to a transient increase in stimulus salience. We model this as a difference of two sigmoids: $a(s) = a_0 + (\overline{a}/(1 + \exp(-4r(t_{s,1} - s))) - (\overline{a}/(1 + \exp(4r(t_{s,2} - s)))$, with parameters $a_0 = -0.04$, $\overline{a} = 0.045$, $t_{s,1} = 0.75$, $t_{s,2} = 1.25$, and $r = 20$: see Fig. 9. Additionally, we take constant noise strength $c(s) \equiv 0.06$ and $\tau_c = \beta = 1$.

For the pure drift-diffusion model, Eq. (47) again gives $\overline{g}_{dd}(s) = \tau_c k a(s)$, and for the

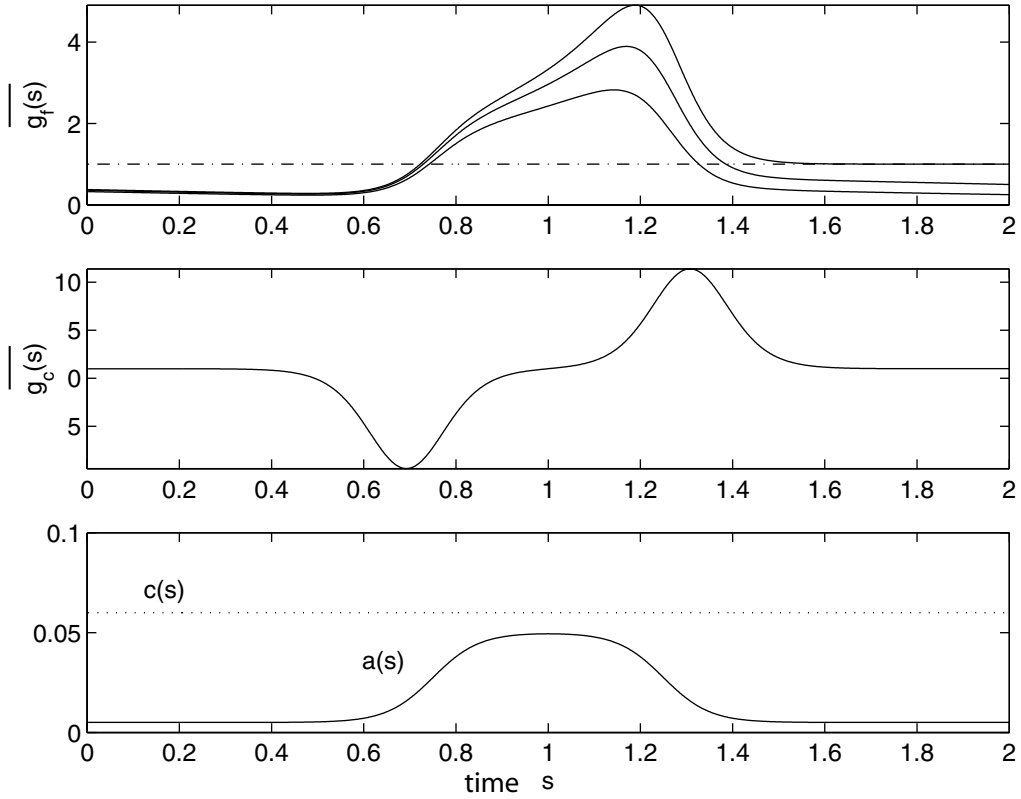


Fig. 9. Optimal gains for pulsed signal strength $a(s)$ (solid line in bottom panel) and constant noise amplitude $c(s) \equiv 0.06$ (dotted line). Top panel: three optimal gain schedules $\overline{g}_f$ for the firing rate model solving (52) (solid curves); the nonoptimal constant gain function $g \equiv 1/\beta$ is shown dot-dashed for reference. Central panel: the unique optimal gain function for the connectionist model, given by Eq. (49).

connectionist and firing rate models, we may use (49) and (52) for the entire time interval of interest since $a(s)$ is strictly positive. The resulting optimal gain trajectories, shown in Fig. 9, yield 70.8% correct responses at interrogation time $T = 2$, compared with 64.9% correct obtained for constant gain $g_f(s) \equiv 1/\beta$ in the firing rate model. Note that the form of the optimal $\overline{g}_c(s)$ illustrates the intuitive explanation given in Sec. 3.3.2: when the signal-to-noise ratio increases, $\overline{g}_c(s)$ *decreases*, suppressing previously integrated information, and vice-versa.

In summary, we have shown in this section that gain schedules yielding optimal performance in (reduced) models of decision tasks depend strongly on the time course of task stimuli as well as the structure of the underlying model, although they all implement matched filters and maximize the signal-to-noise ratio in the difference between activities of neural populations representing competing alternatives. For systems well-described by connectionist models, neural mechanisms may be expected to depress the gain (i.e. strength of inhibitory feedback) below the "balanced" level of $1/\beta$ when stimulus salience is increasing, and enhance it above this level when salience is decreasing. However, for the firing rate model an optimal network can "choose" among a variety of gain schedules of qualitatively different forms. One neurobiological implication of this flexibility is explored in the following section.

4. The Locus Coeruleus Brainstem Area and Optimal Gain Trajectories

Neurons comprising the brainstem nucleus locus coeruleus (LC) emit the neurotransmitter norepinephrine (NE) to targets widely distributed throughout the brain, including cortical areas involved in decision tasks. While NE has disparate and complex effects on different brain regions, a dominant cortical role is believed to be modulation of neuronal gain at both the single cell and population levels [Usher *et al.*, 1999; Servan-Schreiber *et al.*, 1990]. Recordings of cortical neuron responses to stereotyped inputs at various latencies following activation of LC reveal these gain effects: responses to a fixed input are larger (in certain experimental ranges) following LC activation than in control recordings without LC, and this elevated sensitivity decays with a time constant $\tau_{NE} \approx 0.2$ sec [Waterhouse *et al.*, 1998].

Since the firing rate of LC neurons governs NE release rate, we propose the following simple model for cortical gain $g(t)$:

$$\tau_{NE}\, \dot{g}(t) = k_{LC}\, LC(t) - g(t)\,. \tag{59}$$

Here, $LC(t)$ denotes the time-dependent rate of LC firing and k_{LC} is a constant relating this rate to equilibrium values of cortical gain. This model's limitations in describing the underlying biology include the fact that $g(t)$ decays to zero in the absence of LC firing (this could be rectified by adding a constant "gain floor" g_{base}). Nevertheless, it allows us to make an interesting qualitative point in relating recent data on LC firing rates to optimal strategies for the processing of noisy sensory stimuli. Inverting (59) and inserting an optimal gain trajectory yields a prediction for the optimal time course of LC activity:

$$\overline{LC}(t) = \frac{1}{k_{LC}}\,(\tau_{NE}\,\dot{\overline{g}}(t) + \overline{g}(t)). \tag{60}$$

Figure 10(d) shows histograms of LC firing rates recorded from monkeys performing two different psychological tasks: target identification, in which a horizontal or vertical bar must be detected, and the Eriksen flanker task, in which a central cue must be identified while an array of distractors is ignored. Since the second task involves more complex stimulus processing, we assume as in [Brown *et al.*, 2004b] that the onset of stimulus representation in cortical decision areas is more gradual in this than in the target identification task. Specifically, for t greater than the time t_s of stimulus arrival we take $a(t) = \overline{a}(1 - e^{-r(t-t_s)})$ with $r = 50$ (time constant 0.02 sec) for target identification and $r = 10$ (time constant 0.1 sec) for the Eriksen task; also, we set $\overline{a} = 0.06$; and $\tau_c = 0.5$ sec: see Fig. 10(b). Additionally, we assume that t_s follows presentation of the sensory cue by a processing time lag of 0.1 sec (cf. [Aston-Jones *et al.*, 1994]). Optimal gain schedules $\overline{g}_f(t)$ for the firing rate model with these stimuli, computed as in the preceding section, are shown in Fig. 10(a). To produce panel (c), these gain functions were inserted into Eq. (60) to yield corresponding optimal LC firing rates, the discontinuity in $\overline{g}_f(t)$ at stimulus onset having negligible effect. (Also note that assuming a smoother profile for $a(t)$ would eliminate the jump in $LC(t)$.) The similarity between overall form and decay rates of optimal gain functions $\overline{LC}(t)$ and the empirical data of Fig. 10(d) supports the

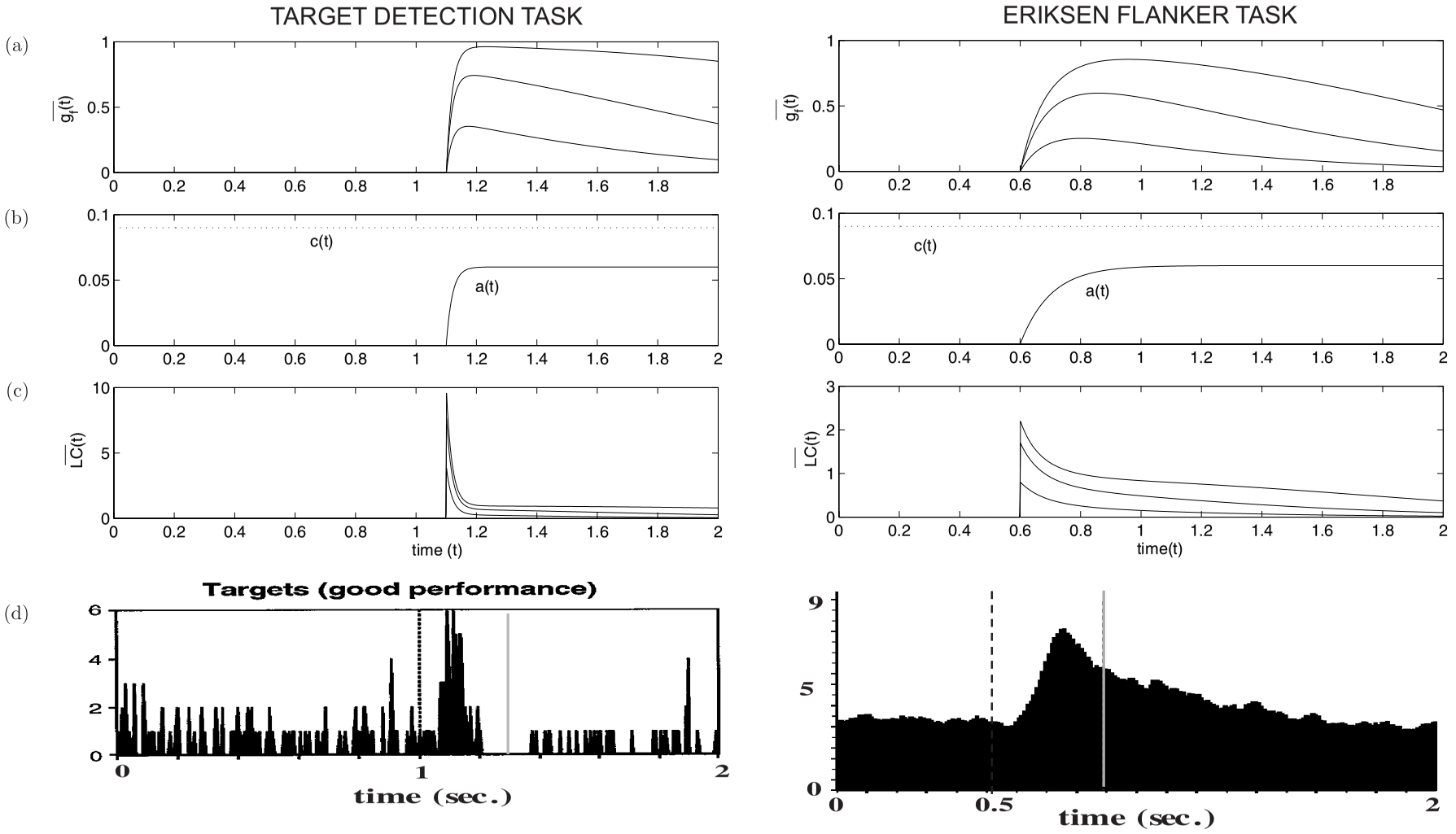


Fig. 10. Comparison of optimal gain theory with empirical data for two psychological tasks. (a) Optimal gain schedules for the firing rate model, for rapid (left) and gradual (right) onset of stimulus $a(t)$ to neural units (with a processing time lag of 0.1 sec following sensory cue), as shown in (b). (c) The corresponding optimal time courses of LC firing rate. (d) Histograms of LC firing rates recorded in the two tasks: (left) the target detection task [Usher *et al.*, 1999] (right) the Eriksen flanker task, with data kindly provided by the authors of [Clayton *et al.*, 2004]. Vertical dashed lines indicate onset of sensory stimuli, and vertical gray (solid) lines indicate mean behavioral reaction time (standard deviations are ≈ 34 and 114 msec for the target detection and Eriksen tasks, respectively).

hypothesis that the LC may affect near-optimal processing of sensory stimuli. This is true even though LC firing rates are not sustained at the initial high values that follow stimulus onset; in fact, both LC firing rate relaxation and NE time constants are compatible with optimal gain schedules.

We note that the optimal gains, and hence $LC(t)$ time courses, are computed assuming prior knowledge of the stimulus $a(t)$ and signal-to-noise ratio $a(t)/c(t)$. If this were the case, LC firing patterns should be well-correlated with stimulus onset. However, experimental data of [Clayton *et al.*, 2004], which involved variable stimulus onset times, indicates tighter correlations with behavioral responses. Here, the function $a(t)$ is perhaps better interpreted as input to motor neurons, the onsets of sensory stimuli having been detected earlier in decision layers. Thus, the most appropriate LC data for use in Fig. 10 would be aligned with transients in firing rates in intermediate processing layers; here we provide data aligned with sensory stimuli as the closest available surrogate. Explicit models of multilayer decision/response dynamics with variable gain are studied in [Brown *et al.*, 2004a].

5. Discussion and Conclusions

In this paper we explicitly compute optimal gain trajectories for one-dimensional, linearized reductions of simplified models for competing neural groups involved in decisions between two alternatives. We also demonstrate via simulations that such reductions provide good approximations for the reaction time and error rate statistics of the nonlinear two-dimensional connectionist and firing rate models from which they were derived.

We first show that the nonlinear connectionist and firing rate models are equivalent, under suitable variable and parameter coordinate changes. We then develop a piecewise-linear approximation to the canonical sigmoidal activation or firing rate function. The resulting two-dimensional piecewise-linear SDEs (15)–(16) introduced in Sec. 2.4 form a midpoint in our simplification process. This system can be easily solved on each of nine "tiles" forming its phase plane, but solutions must be assembled by matching constants of integration. To illustrate this, we focus on two specific cases in Sec. 2.5.1, motivated by the moving dots' paradigm [Britten *et al.*, 1993; Shadlen & Newsome, 2001; Gold & Shadlen, 2002], that correspond to differing stimulus presentation conditions and rely on different neural mechanisms to implement transient effective gain values.

In Case 1, the development of salience (i.e. $a_1 \neq a_2$), in sensory stimuli at time t_s is not accompanied by large changes in the stimulus magnitudes; in fact the summed magnitude is unchanged. This mild stimulus onset is insufficient to move solutions between tiles, so variations in gain must result from modulation of the gain of the neural activation function itself, presumably via influence of other brain areas such as the locus coeruleus. However, in Case 2, the appearance of salience is accompanied by large changes in stimulus magnitude, either due to properties of the stimulus itself or due to additive biases that shift the activation function to the left, as has been proposed in connectionist models that address the effects of attention [Mozer, 1988; Cohen *et al.*, 1992]. In this case, no external modulation of gain is required, since the decision dynamics themselves move the system between regions of the activation function where desired sensitivities (and hence gains) are achieved. The possibility that neural systems are tuned so that the presence of target stimuli causes solutions to move into sensitive regions of their activation functions has been previously suggested in behavioral neuroscience [Servan-Schreiber *et al.*, 1990]; here we reformulate this idea in terms of optimal signal processing.

We end by showing that the (nonunique) optimal gain schedules for the firing rate model include time courses that are consistent with release of norepinephrine due to transient increases in the activity of neurons in locus coeruleus.

The external modification of gain considered in Case 1 assumes prior knowledge of the time course of the absolute values of sensory inputs $a_j(t)$, the task of the decision maker being merely to identify their signs. In [Brown *et al.*, 2004a] the more general case in which this information is not available is treated, and strategies must additionally include a mechanism for detecting increases in signal-to-noise ratio of sensory inputs.

Acknowledgments

This work was partially supported by DoE grant DE-FG02-95ER25238 and PHS grants MH58480 and MH62196 (Cognitive and Neural Mechanisms of Conflict and Control, Silvio M. Conte Center). E. Brown was supported under a National Science Foundation Graduate Fellowship and a Burroughs–Wellcome Training Grant in Biological Dynamics.

The authors thank Josh Gold and Jaime Cisternas for useful contributions and discussions, as well as Ed Clayton and Gary Aston-Jones for providing the data of Fig. 10 and for their insights into the role of the LC in modulating decisions.

References

Abbott, L. [1991] "Firing-rate models for neural populations," in *Neural Networks*: *From Biology to High-Energy Physics*, eds. Benhar, O., Bosio, C., Del Giudice, P. & Tabat, E. (ETS Editrice, Pisa), pp. 179–196.

Amit, D. & Tsodyks, M. [1991] "Quantitative study of attractor neural network retrieving at low spike rates: I. Substrate–spikes, rates, and neuronal gain," *Network* **2**, 259–273.

Anderson, J. [1990] *The Adaptive Character of Thought* (Lawrence Erlbaum, Hillsdale, NJ).

Arnold, L. [1974] *Stochastic Differential Equations* (John Wiley, NY).

Arnold, L. [1998] *Random Dynamical Systems* (Springer, Heidelberg).

Aston-Jones, G., Rajkowski, J., Kubiak, P. & Alexinsky, T. [1994] "Locus coeruleus neurons in the monkey are selectively activated by attended stimuli in a vigilance task," *J. Neurosci.* **14**, 4467–4480.

Bialek, W., Rieke, F., de Reuter van Steveninck, R. & Warland, D. [1991] "Reading a neural code," *Science* **252**, 1854–1857.

Bogacz, R., Brown, E., Moehlis, J., Hu, P., Holmes, P. & Cohen, J. D. [2004] "The physics of optimal decision making: A formal analysis of models of performance in two alternative forced choice tasks," *Psych. Rev.*, in review.

Boxler, P. [1991] "How to construct stochastic center manifolds on the level of vector fields," in *Lyapunov Exponents*, eds. Arnold, L., Crauel, H. & Eckmann, J.-P., Lecture Notes in Mathematics, Vol. 1486 (Springer, Heidelberg), pp. 141–158.

Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, J. A. [1993] "Responses of neurons in macaque MT to stochastic motion signals," *Vis. Neurosci.* **10**, 1157–1169.

Brown, E. & Holmes, P. [2001] "Modeling a simple choice task: stochastic dynamics of mutually inhibitory neural groups," *Stochast. Dyn.* **1**, 159–191.

Brown, E., Gilzenrat, M. S. & Cohen, J. D. [2004a] "The locus coeruleus, adpative gain, and the optimization of simple decision tasks," Technical Report #04–02, Center for the Study of Mind, Brain, and Behavior, Princeton University.

Brown, E., Moehlis, J., Holmes, P., Clayton, E., Rajkowski, J. & Aston-Jones, G. [2004b] "The influence of spike rate and stimulus duration on noradrenergic neurons," *J. Comput. Neurosci* **17**, 5–21.

Brunel, N., Chance, F., Fourcaud, N. & Abbott, L. F. [2001] "Effects of synaptic noise and filtering on the frequency response of spiking neurons," *Phys. Rev. Lett.* **86**, 2186–2189.

Chance, F. S., Abbott, L. F. & Reyes, A. D. [2002] "Gain modulation from background synaptic input," *Neuron* **35**, 773–782.

Cho, R., Nystrom, L., Brown, E., Jones, A., Braver, T., Holmes, P. & Cohen, J. D. [2002] "Mechanisms underlying performance dependencies on stimulus history in a two-alternative forced choice task," *Cogn. Affect. Behav. Neurosci.* **2**, 283–299.

Clayton, E., Rajkowski, J., Cohen, J. D. & Aston-Jones, G. [2004] "Decision-related activation of monkey locus coeruleus neurons in a forced choice task," under preparation.

Cohen, J. D., Dunbar, K. & McClelland, J. L. [1990] "On the control of automatic processes: A parallel distributed processing model of the Stroop effect," *Psychol. Rev.* **97**, 332–361.

Cohen, J. D., Servan-Schreiber, D. & McClelland, J. L. [1992] "A parallel distributed processing approach to automaticity," *Amer. J. Psychol.* **105**, 239–269.

Cohen, J. D. & Huston, T. A. [1994] "Progress in the use of interactive models for understanding attention and performance," in *Attention and Performance XV*, eds. Umilta, C. & Moscovitch, M. (MIT Press, Cambridge), pp. 453–476.

Ermentrout, G. B. [1994] "Reduction of conductance-based models with slow synapses to neural nets," *Neur. Comput.* **6**, 679–695.

Fairhall, A., Lewen, G., Bialek, W. & de Ruyter van Steveninck, R. [2001] "Effciency and ambiguity in an adaptive neural code," *Nature* **412**, 787–792.

Gardiner, C. W. [1985] *Handbook of Stochastic Methods*, 2nd edition (Springer, NY).

Gerstner, W. & Kistler, W. [2002] *Spiking Neuron Models* (Cambridge University Press, Cambridge, UK).

Gilzenrat, M. S., Holmes, B. D., Rajkowski, J., Aston-Jones, G. & Cohen, J. D. [2002] "Simplified dynamics in a model of noradrenergic modulation of cognitive performance," *Neural Networks* **15**, 647–663.

Gold, J. I. & Shadlen, M. N. [2001] "Neural computations that underlie decisions about sensory stimuli," *Trends Cogn. Sci.* **5**, 10–16.

Gold, J. I. & Shadlen, M. N. [2002] "Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward," *Neuron* **36**, 299–308.

Grossberg, S. [1988] "Nonlinear neural networks: Principles, mechanisms, and architectures," *Neural Networks* **1**, 17–61.

Guckenheimer, J. & Holmes, P. J. [1983] *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields* (Springer-Verlag, NY).

Hertz, J., Krough, A. & Palmer, R. [1991] *Introduction to the Theory of Neural Computation* (Perseus Book Group, NY).

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

Huk, A., Palmer, J. & Shadlen, M. [2002] "Temporal integration of motion energy underlies perceptual decisions and response times," *Annual Society for Neuroscience Meeting*, Orlando, FL, Nov 2–7, 2002, Abstract No. 353.5.

Knobloch, E. & Weisenfeld, K. A. [1983] "Bifurcations in fluctuating systems: The center manifold approach," *J. Stat. Phys.* **33**, 611–637.

Laming, D. R. J. [1968] *Information Theory of Choice-Reaction Times* (Academic Press, NY).

Lehmann, E. L. [1959] *Testing Statistical Hypotheses* (John Wiley, NY).

McClelland, J. L. [1979] "On the time relations of mental processes: An examination of systems of processes in cascade," *Psychol. Rev.* **86**, 287–330.

Mozer, M. [1998] "A connectionist model of selective attention in visual perception," in *Proc. Tenth Ann. Conf. Cognitive Science Society* (Erlbaum, Hillsdale, NJ), pp. 195–201.

Omurtag, A., Kaplan, E., Knight, B. W. & Sirovich, L. [2000] "A population approach to cortical dynamics with an application to orientation tuning," *Network* **11**, 247–260.

Papoulis, A. [1977] *Signal Analysis* (McGraw-Hill, NY).

Platt, M. L. & Glimcher, P. W. [2001] "Neural correlates of decision variable in parietal cortex," *Nature* **400**, 233–238.

Ratcliff, R. [1978] "A theory of memory retrieval," *Psych. Rev.* **85**, 59–108.

Ratcliff, R., Van Zandt, T. & McKoon, G. [1999] "Connectionist and diffusion models of reaction time," *Psych. Rev.* **106**, 261–300.

Roitman, J. & Shadlen, M. [2002] "Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task," *J. Neurosci.* **22**, 9475–9489.

Schall, J. D. [2001] "Neural basis of deciding, choosing, and acting," *Nature Rev.: Neurosci.* **2**, 33–42.

Schall, J., Stuphorn, V. & Brown, J. [2002] "Monitoring and control of action by the frontal lobes," *Neuron* **36**, 309–322.

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

Shadlen, M. N. & Newsome, W. T. [2001] "Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey," *J. Neurophysiol.* **86**, 1916–1936.

Shelley, M. & McLaughlin, D. [2002] "Coarse-grained reduction and analysis of a network model of cortical response. I. drifting grating stimuli," *J. Comput. Neurosci.* **12**, 97–122.

Shin, J., Koch, C. & Douglas, R. [1999] "Adaptive neural coding dependent on the time varying statistics of the somatic input current," *Neural Comput.* **11**, 1083–1913.

Smith, P. L. & Ratcliff, R. [2004] "Psychology and neurobiology of simple decisions," *Trends in Neurosci.* **27**, 161–168.

Stone, M. [1960] "Models for choice-reaction time," *Psychometrika* **25**, 251–260.

Usher, M., Cohen, J. D., Servan-Schreiber, D., Rajkowsky, J. & Aston-Jones, G. [1999] "The role of locus coeruleus in the regulation of cognitive performance," *Science* **283**, 549–554.

Usher, M. & McClelland, J. L. [2001] "On the time course of perceptual choice: The leaky competing accumulator model," *Psych. Rev.* **108**, 550–592.

von Neumann, J. [1958] *The Computer and the Brain* (Yale University Press, New Haven, CT); 2nd edition [2000], with a foreword by Paul and Patricia Churchland.

Wald, A. [1947] *Sequential Analysis* (John Wiley, NY).

Wang, X.-J. [2002] "Probabilistic decision making by slow reverberation in cortical circuits," *Neuron* **36**, 955–968.

Waterhouse, B., Moises, H. & Woodward, D. [1998] "Phasic activation of the locus coeruleus enhances responses of primary sensory cortical neurons to peripheral receptive field stimulation," *Brain Res.* **790**, 33–44.

Wickelgren, W. A. [1977] "Speed-accuracy tradeoff and information processing dynamics," *Acta Psychol.* **41**, 67–85.

Wilson, H. & Cowan, J. [1972] "Excitatory and inhibitory interactions in localized populations of model neurons," *Biophys. J.* **12**, 1–24.