

Optimization of decision making in multilayer networks: The role of Locus Coeruleus

Eric Shea-Brown^{1,3,4}, Mark S. Gilzenrat^{2,3}, and Jonathan D. Cohen^{2,3}

¹ Program in Applied and Computational Mathematics,

² Department of Psychology,

³ Program in Neuroscience, Princeton University, Princeton, NJ 08544, U.S.A.

⁴ Courant Institute and Center for Neural Science,
New York University, New York, NY 10012, U.S.A.

Abstract

Previous theoretical work has shown that a single layer neural network can implement the optimal decision process for simple, two alternative forced choice (2AFC) tasks. However, it is likely that the mammalian brain is comprised of multilayer networks, raising the question of whether and how optimal performance can be approximated in such an architecture. Here, we present theoretical work suggesting that the noradrenergic nucleus locus coeruleus (LC) may help optimize 2AFC decision making in the brain. This is based on the observations that neurons of the LC selectively fire following the presentation of salient stimuli in decision tasks, and that the corresponding release of norepinephrine can transiently increase the responsivity, or *gain*, of cortical processing units. We describe computational simulations that investigate the role of such gain changes in optimizing performance of 2AFC decision making. In the tasks we model, no prior cueing or knowledge of stimulus onset time is assumed.

Performance is assessed in terms of the rate of correct responses over time ('reward rate'). We first present the results of a single layer model that accumulates (integrates) sensory input and implements the decision process as a threshold crossing. Gain transients, representing the modulatory effect of the LC, are driven by separate threshold crossings in this layer. We optimize over all free parameters to determine the maximum reward rate achievable by this model, and compare it to the maximum reward rate when gain is held fixed. We find that the dynamic gain mechanism yields no improvement in reward for this single layer model.

We then examine a two layer model, in which competing sensory accumulators in the first layer (capable of implementing the task relevant decision) pass activity to response accumulators in a second layer. Again, we compare a version in which threshold crossing in the first (decision) layer elicits an LC response (and a concomitant increase in gain) with a fixed gain version of the model. Here, we find that gain transients modeling the LC phasic response yield an improvement in reward rate of 12 – 24%. Furthermore, we show that the timing characteristics of these gain transients agree with observations concerning LC firing patterns reported in recent experimental studies. This provides converging evidence for the hypothesis that the LC optimizes processes underlying 2AFC decision making in multilayer networks.

1 Introduction and background

Implementation of optimal decision making in neural networks

Research in cognitive neuroscience has begun to characterize the neural mechanisms underlying decision making in simple, 2AFC tasks (e.g., (Schall and Thompson, 1999; Schall and Thompson, 1996; Gold and Shadlen, 2002; Hanes and Schall, 1996; Shadlen and Newsome, 2001)). This work has produced evidence that both the dynamics of neural activity and behavioral performance in such tasks can be described accurately by a simple mathematical model, often referred to as the drift diffusion model (DDM (Gold and Shadlen, 2002; Schall and Thompson, 1999; Ratcliff, Thapar, and McKoon, 2003)). The DDM describes decision processes in terms of simple accumulators that integrate signals favoring each of the two choices, and respond when the difference between these exceeds a threshold value (cf. (Stone, 1960; Laming, 1968; Ratcliff, 1978; Smith and Ratcliff, 2004)). Under certain constraints, the DDM can be shown to implement a continuous version of the Sequential Probability Ratio Test (SPRT), which is known to be the optimal algorithm for 2AFC decision making (Laming, 1968; Wald, 1947; Lehmann, 1959; Stone, 1960). Furthermore, recent analyses have shown that the DDM represents a good approximation of single layer neural network models that accurately simulate performance in 2AFC tasks (Bogacz et al., 2006; Usher and McClelland, 2001). Typically, these networks are comprised of sets of competing units, each of which accumulates evidence in favor of a particular alternative, and the competition among which implements the decision process.

From the forgoing observations, it can be inferred that single layer neural networks are optimal for 2AFC decision processing. At the same time, it is evident that such mechanisms are imbedded in larger networks in the brain. The simplest reason for this is that the same integrated sensory information must

be able to be flexibly translated into different responses (e.g., finger movement or articulatory codes) in different tasks (Schall, 2003). Moreover, behavioral and physiological evidence ((Schall, 2003; Reddi, 2001) and references therein) suggests that at least two stages of processing contribute significantly to variability in response times.

This presents a challenge for optimal performance. On the one hand, as we have noted above, a single layer network implements the optimal decision process. On the other hand, the layer implementing this process for a given task may be at least one layer removed from the response mechanism. It would be inefficient if a decision unit in the layer integrating information relevant to the current task crossed threshold, but then had to drive subsequent accumulator processes - which introduce additional noise and require additional integration time - before a behavioral response can be elicited. In other words, there is a fundamental tradeoff between the complexity of a multilayer system that can support a wide range of decision processes (and the flexibility of behavior that this affords), and the efficiency of a simpler, single layer system (that is, the optimality of function that this affords).

The inefficiency of multilayer processing can be ameliorated if, at the time a unit in the task-relevant decision layer crosses threshold, a signal is issued ensuring that this information rapidly and directly influences the behavioral response. Recent evidence suggests that phasic release of norepinephrine from the locus coeruleus following the decision may serve this function.

The Locus Coeruleus

Neurons of the nucleus locus coeruleus (LC) selectively fire following the presentation of salient stimuli in decision tasks (Usher et al., 1999; Aston-Jones et al., 1994; Clayton et al., 2004). This produces widespread release of norepinephrine throughout the neocortex, that is thought to potentiate the response of target neurons to their current input (Waterhouse, Moises, and Woodward, 1998). Thus, the LC can provide transient facilitation in processing throughout the cortex, time-locked to the presence of behaviorally salient information. This facilitation has been modeled as a change in the gain of processing units representing populations of neurons involved in the decision making task (Servan-Schreiber, Printz, and Cohen, 1990; Usher et al., 1999; Gilzenrat et al., 2002; Usher and Davelaar, 2002). Simulation studies based on this model have shown that transient increases in gain locked to the decision process in a simple signal detection task can reduce error rates without increasing reaction time, thus effecting an improvement in behavioral performance, and not merely a shift in the speed-accuracy tradeoff (Usher et al., 1999; Gilzenrat et al., 2002; Usher and Davelaar, 2002). This corresponds to the empirical observation that phasic LC responses are associated with epochs of

maximal performance (Aston-Jones et al., 1994). These findings suggest that the LC may play an important role in optimizing processing. However, previous simulation studies have parameterized models to fit empirical data, and have not systematically examined the question of whether there is a principled benefit of dynamic adjustments in gain for optimizing performance, as compared to an optimized fixed value of gain. To answer this question it is necessary to compare the best-possible levels of performance across a class of dynamic vs. fixed gain mechanisms, and over a sufficiently wide range of free parameters. This is the approach we take here.

In accord with previous work addressing optimal performance in 2AFC decision tasks (Gold and Shadlen, 2002; Bogacz et al., 2006), we use reward rate as our metric of task performance; that is, the average amount of reward obtained per unit time. Note that reward rate is a function of both speed and accuracy.

As discussed above, previous studies of 2AFC decision tasks (Smith and Ratcliff, 2004; Stone, 1960) have shown that, under certain constraints, the DDM corresponds to a continuous version of the SPRT (the optimal decision algorithm), and that this can be implemented in a simple single layer neural network. The constraints require that the timing of stimuli is perfectly known to the model, and that stimuli remain constant in strength when presented. These results have been extended to show that given the same constraints of perfect knowledge of stimulus timing, but allowing for varying stimulus strength (i.e. waxing and waning perceptual salience during the course of a decision trial), the same network can recover optimal performance if its gain is dynamically regulated by an LC-like mechanism (Brown et al., 2005). However, for the case in which the timing of task stimuli is unknown (i.e., there is no cue or obvious perceptual change that precedes or accompanies the appearance of the stimuli), it is no longer possible to apply the SPRT and an entirely new analysis is required. Here, we consider this more general case.

Our first studies implement the same single layer mutually inhibitory network studied in (Bogacz et al., 2006; Usher and McClelland, 2001; Brown and Holmes, 2001; Brown et al., 2005). We compare the performance of this model with two gain mechanisms. The first is a fixed gain mechanism in which a constant value of gain is used throughout the course of each decision trial and across trials. This implementation provides our control condition. The second is a dynamic gain mechanism intended to simulate the effects of the LC, in which a stepped change in gain may occur within a decision trial. We assume that this gain transient is triggered by the accumulation of sensory inputs past some critical value. We separately optimize parameters for the fixed gain and dynamic gain models, to compare the advantage in reward rate, if any, afforded by an LC-like transient gain mechanism operating within a single decision layer.

The single layer model makes contact with optimal architectures investigated in (Brown et al., 2005; Bogacz et al., 2006). However, as discussed above, decision mechanisms relevant to a particular task are almost certainly embedded in a more complex processing architecture within real brains. This raises the question of how dynamic adjustments of gain may impact the performance of networks composed of more than a single processing layer.

To address this question, we also study the simplest case of a multilayer system: a two layer model. In this model, the first layer implements the same decision making mechanisms as the single layer model. However, the output of this layer then feeds to a second, response layer that simulates the behavioral response. We assume that processing occurs on roughly similar timescales in both layers, an important point revisited in the Discussion. As before, we either restrict this model to have fixed gain independently set for each layer, or allow gain to vary simultaneously across both layers, in this case driven by the decision making process in the first layer. We again separately optimize the parameters of each model implementation to determine the maximum reward rate achievable with fixed vs. dynamic gain.

An optimized decision model equipped with LC-mediated gain transients will clearly perform no worse than an optimized model for which gain values are held fixed (since, in the limit, a dynamic gain model can be parameterized so that changes in gain never occur). However, it is not immediately apparent that a dynamic gain model will perform any better, at least as implemented by the LC. There are several reasons for this. One is that changes in gain via the LC are subject to latencies (i.e., conduction time, delay in neuromodulatory effects at target sites, etc.). For example, it has been shown (Waterhouse, Moises, and Woodward, 1998) that potentiation of cortical neurons to sensory inputs follows stimulation of the LC by more than 100 ms, reaching a plateau of maximum effect at 200-300 ms. In order to similarly constrain our model, we employ a fixed delay in gain effects of 150 ms. This delay, together with the latency from stimulus onset to onset of the gain transient, mean that gain changes occur appreciably after the presentation of stimuli, possibly minimizing their benefit.

A further issue along these lines arises from our assumption that there is no separate information about the timing of stimulus onsets, so that gain transients must be driven by accumulation of task inputs on a trial-to-trial basis. Even if a dynamic gain mechanism provides significant benefits in performance when appropriately triggered, gain transients elicited inappropriately (due to noise, for example) may lead to erroneous responding that negates the benefits derived from gain transients on correct trials. These issues raise a more general problem of “circularity” in determining when to initiate a transient increase in gain: Is the amount of information that must be integrated to determine that a change in gain is warranted already sufficient to make a decision about the

identity of the stimulus itself, obviating the need for an increase in gain? That is, is the decision-making process sufficiently unresolved at the time the gain transient is triggered that, given its inherent latencies, there is still time for a gain change to produce facilitative effects on performance? We will consider this problem in light of our results in the Discussion.

We find the following answers to the questions just introduced. In single layer systems, dynamic adjustments in gain do not enhance reward rate beyond the best values achievable with fixed gain values. For two layer decision models, however, LC mediated gain transients yield moderate performance enhancements relative to their fixed gain counterparts. Section 2 introduces the decision models and the performance measure of reward rate. In Section 3 we describe the method, central to the present analysis, of comparing optimal performance of models that used fixed versus dynamic gain mechanisms. In Section 4, we present the results of numerical reward rate optimization for the single and two layer models. Finally, an Appendix shows that our main results persist under several modifications to the decision model, and describes mathematical details.

2 Models and reductions

The single layer model

The first decision model we investigate, as shown in Figure 1, consists of a pair of mutually inhibitory neural units, each of which is meant to represent a population of neurons involved in the decision making task¹. Each unit receives input corresponding to the level of sensory evidence for one of the two response alternatives. The activity of the two units evolves according to ‘rate equations’ of the general form derived in (Wilson and Cowan, 1972), and widely applied and analyzed (e.g. in (Hopfield, 1984; Abbott, 1991; Gerstner and Kistler, 2002)). This model is a variant of the competing leaky accumulator model introduced by (Usher and McClelland, 2001) and further studied in (Brown and Holmes, 2001; Bogacz et al., 2006). The firing rates y_1 and y_2 of the two decision units evolve according to:

$$\tau \frac{dy_1}{dt} = -y_1 + f_{\tilde{g}_y(t)}(-\beta y_2 + ka_1(t)) + \tilde{g}_y(t) \frac{kc\sqrt{\tau}}{\sqrt{2}} \eta^1(t), \quad (1)$$

$$\tau \frac{dy_2}{dt} = -y_2 + f_{\tilde{g}_y(t)}(-\beta y_1 + ka_2(t)) + \tilde{g}_y(t) \frac{kc\sqrt{\tau}}{\sqrt{2}} \eta^2(t). \quad (2)$$

¹Each unit is described by an activity, which represents the instantaneous firing rate of the corresponding population. See (Machens, Romo, and Brody, 2005; Wong and Wang, 2006) for comparisons between the dynamics of spiking neural populations and dynamics on the simpler level of interacting units.

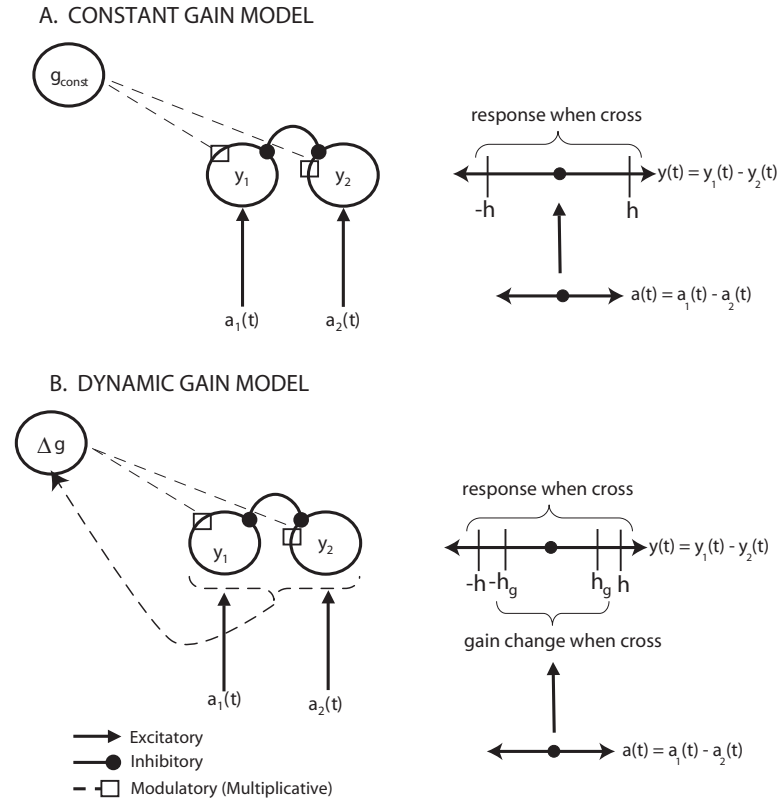


Figure 1: Single layer models of the decision task. We study the difference in firing rates y_1 and y_2 , reducing the system to motion on a line (right). A. The fixed gain case, in which gain is held constant throughout the trial. B. The dynamic gain case. Gain is held at a fixed value until $y(t)$ crosses a ‘gain threshold’ $\pm h_g$, at which time gain is increased to some new optimized value (after a fixed delay). In both cases, when $|y|$ exceeds h , a response is made. Note that although the gain threshold is shown within the bounds of the response threshold, the model was not constrained to this arrangement.

Here, the parameter β sets the strength of mutual inhibition between the units, which scales with the firing rate of the opposing unit. Sensory inputs corresponding to alternatives ‘1’ and ‘2’ are represented by $ka_1(t)$ and $ka_2(t)$, where k is a factor scaling these inputs. To model alternative “1,” we take $a_1(t) \geq a_2(t)$, and vice-versa for alternative “2.” These inputs, as well as incoming mutual inhibition, enter as arguments to the sigmoidal ‘activation’ (or ‘frequency-current’) function $f_{g(t)}(\cdot)$. The subscript in $f_{g(t)}(\cdot)$ indicates dependence on the time-varying gain, or sensitivity, $g(t)$ of the activation function. Specifically, the gain determines the maximum slope, as for the common choice $f_{g(t)}(x) = 1/2 [1 + \tanh(2g(t)(x - b))]$.

Fluctuations in the inputs are represented by independent white noise processes $\eta^1(t)$ and $\eta^2(t)$ (of unit variance). The noise strength c sets the amplitude of these noise terms, along with the scale factor k . We assume that 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).

We next make our first simplifying assumption: that the inputs to $f_{g(t)}(\cdot)$ remain in a range where this function may be approximated by its linearization, i.e. $f_{g(t)}(x) \approx 1/2 + g(t)(x - b)$. Thus, we assume our network to be biased such that the model units operate within the linear range (i.e. the most responsive portion) of their activation functions during the decision task (Cohen, Dunbar, and McClelland, 1990). We note that the equivalence of such two-unit firing rate models to the optimal SPRT algorithm (see Introduction) has only been demonstrated for such linearized networks ((Bogacz et al., 2006), cf. (Brown et al., 2005)).

Next, we define the firing rate difference $y = y_1 - y_2$ (corresponding to the decision variable in the DDM), as well as the stimulus salience $a = a_1 - a_2$ (corresponding to the drift rate in the DDM). We then study the (linearized) model involving only these differences:

$$\tau \frac{dy}{dt} = -\tilde{y} + g_y \beta y + g_y k a + g_y k c \sqrt{\tau} \eta(t) . \quad (3)$$

This equation is obtained by subtracting Eqns. (1) and (2), using the above linearized activation function. Here, $\eta(t)$ is again a unit variance white noise process. Further, ka sets the level of input associated with sensory ‘evidence’ of task stimuli, with positive values favoring alternative 1 and negative values alternative 2. Finally, we take initial conditions $y(0) = 0$ at the beginning of the trial, effectively assuming that firing rates have returned to baseline by that time.

In adopting the reduced model (3) from this point forward, we make our second simplifying assumption: that decisions are determined entirely by the relative (i.e., subtracted) firing rates of the two units. Besides making the

optimization problems we undertake below more tractable, this simplifying choice has empirical grounding. First, neuroimaging studies of perceptual decision-making in human striate and extrastriate visual cortex (e.g. (Simoncelli and Heeger, 1998; Heeger et al., 1999)) suggest that differences in the activity of competing neural populations are critical in driving decision behaviors. Furthermore, drift-diffusion models of two-alternative choice tasks have demonstrated that processes diffusing in a single dimension (representing the difference in evidence accumulated for one vs. the other alternative) capture response time distributions and accuracy rates in these tasks (Ratcliff, Van Zandt, and McKoon, 1999; Smith and Ratcliff, 2004). Finally, we note that previous modeling studies show that the difference variable y alone characterizes the original two-dimensional model if solutions collapse to a one-dimensional ‘decision manifold’ (Usher and McClelland, 2001; Brown and Holmes, 2001; Bogacz et al., 2006; Brown et al., 2005), although these results would need to be extended to apply to the dynamic gain models studied here.

For the transient gain case, we make the additional assumption that g_y undergoes a stepped change at a delay $t_{NE} = 150 \text{ ms}$ following the time T_g at which y first crosses one of the ‘gain thresholds’ $\pm h_g$. Specifically, we assume that the gain values jump from g_y^{pre} to g_y^{post} at the time $T_g + \tau_{NE}$. This mechanism for adjusting gain generalizes the approach of (Usher et al., 1999; Gilzenrat et al., 2002; Usher and Davelaar, 2002) from target identification tasks (in which gain transients are driven only by the firing rate of the unit responsive to target stimuli) to the two-alternative forced choice task (in which gain transients may be driven by the unit responsive to either of the two possible stimuli). Nevertheless, in modeling changes in gain via a delayed step function, rather than a curvilinear trajectory as in (Usher et al., 1999; Gilzenrat et al., 2002; Brown et al., 2005), we simplify the optimization problem. Finally, we assume that norepinephrine levels have decayed to their baseline levels at the time each trial begins, so that $g_y = g_y^{pre}$.

Response thresholds are placed at $\pm h$. Specifically, a response corresponding to alternative 1 (i.e., $y_1 > y_2$) is made if $y(t)$ first crosses the threshold at $+h$, and likewise for alternative 2 and $-h$. We call the first time at which either of these thresholds is crossed the response time, T . The sequence of events leading to a response is illustrated schematically in Fig. 2.

To determine optimal performance of the models under different architectures and assumptions about the time dependence of gain, we will allow the gain g_y and thresholds h and h_g to vary freely, as specified in greater detail below. We can eliminate parameters β and k in Eqn. (3) (i.e., set them to one) since compensatory values for these parameters can be found for any values of g_y , h and h_g ; that is, setting these values equal to one has no effect on the result of the optimization problem that we solve (see Appendix 2). This yields

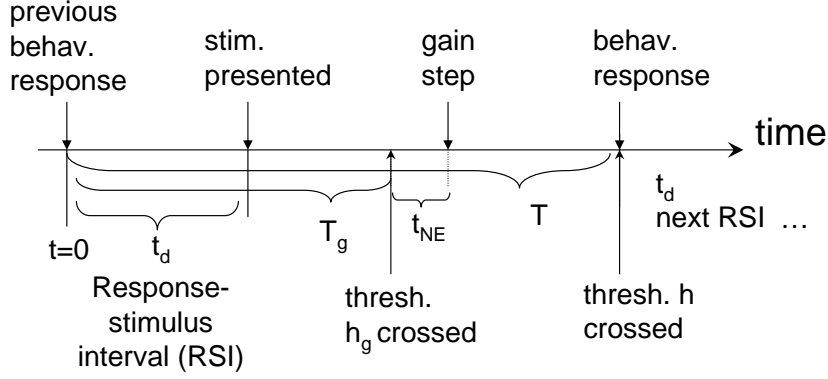


Figure 2: Schematic of events leading to a response. Following a behavioral response on the ‘previous’ trial, model time t is reset to $t = 0$, initial condition $y(0) = 0$ is applied, and firing rates begin evolving under Eqns. (4) or (9-10). The stimulus is presented following a (randomly distributed) time t_d (response-stimulus interval, RSI). For dynamic gain models, the gain undergoes a stepped change at a delay t_{NE} after the firing rate difference y first crosses one of the gain thresholds $\pm h_g$. Finally, the behavioral response is made at time T when the decision variable crosses threshold $\pm h$. Although the ordering of t_d , T_g , and T displayed here is typical, it is not enforced: for example, on some trials gain threshold crossings at T_g could occur before stimulus presentations at t_d .

the simplified system

$$\tau \frac{dy}{dt} = -y + g_y y + g_y a + g_y c \sqrt{\tau} \eta(t) \quad (4)$$

that we will use from this point forward.

The two-layer decision model

In this model we add an additional ‘response layer,’ the activity of which simulates the behavioral response (see Fig. 3), rather than having responses determined by the same layer that integrates sensory inputs (as for the single layer model). This second layer receives inputs from the first layer as well as independent perturbations from noise. The dynamics of the response layer, including mutual inhibition, are similar to those of the first layer:

$$\tau \frac{dz_1}{dt} = -z_1 + f_{g_y(t)} (-\beta z_2 + w y_1(t)) + g_z(t) \frac{kc\sqrt{\tau}}{\sqrt{2}} \eta^3(t), \quad (5)$$

$$\tau \frac{dz_2}{dt} = -z_2 + f_{g_y(t)} (-\beta z_1 + w y_2(t)) + g_z(t) \frac{kc\sqrt{\tau}}{\sqrt{2}} \eta^4(t). \quad (6)$$

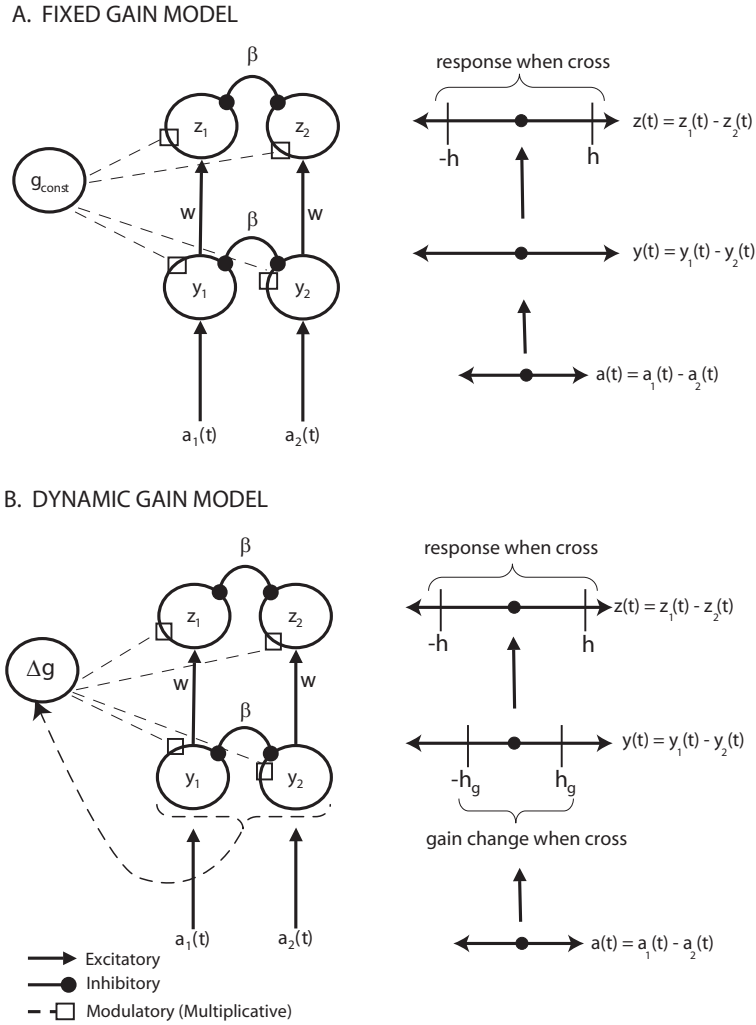


Figure 3: Two layer models of the decision task, which include a second ‘response layer.’ (Left), network representation; (Right), reduction to difference variables. A. The fixed gain case, where gain is held constant throughout the trial. B. The dynamic gain case, in which stepped changes in gain are triggered (a fixed delay after) $y(t)$ crosses ‘gain thresholds’ $\pm h_g$. In both cases, when $|z|$ exceeds h , a response is made.

Here, the weight w scales the connection between the layers. The similar structure of the equations describing each layer is an important modeling assumption whose consequences we will revisit in the Discussion.

As in the one layer case, we next define the firing rate difference $z = z_1 - z_2$ and study the linearized model:

$$\tau \frac{dz}{dt} = -z + g_z \beta z + w g_z y + g_z k c \sqrt{\tau} \eta^1(t) \quad (7)$$

$$\tau \frac{dy}{dt} = -y + g_y \beta y + g_y k a + g_y k c \sqrt{\tau} \eta^2(t) \quad (8)$$

Also as above, a stepped change in gain occurs at a delay τ_{NE} following the first time that y crosses one of the thresholds $\pm h_g$. At this time, gain changes from g_y^{pre} to g_y^{post} in the first layer, and from g_z^{pre} to g_z^{post} in the second layer. Responses are based on activity in the second layer: if $z(t)$ first crosses the threshold at $+h$, a response in favor of alternative 1 is made, and likewise for $-h$ and alternative 2.

As above, the gain and threshold parameters are freely varied to optimize the performance of the two-layer model for fixed values of the remaining parameters. The values of the β and k parameters were again set equal to one: as for the one-layer case, the optimal performance of the two-layer model is independent of the values of these parameters, since the remaining parameters of the model can be rescaled to compensate (see Appendix 2). However, for the two layer model, the ratio of β and w does affect optimal performance. We chose $w = 1$ for simplicity, thereby setting the strength of connections between all units in the model to be identical (see Appendix 2). Hence, we have the simplified model

$$\tau \frac{dz}{dt} = -z + g_z z + g_z y + g_z c \sqrt{\tau} \eta^1(t) \quad (9)$$

$$\tau \frac{dy}{dt} = -y + g_y y + g_y a + g_y c \sqrt{\tau} \eta^2(t) \quad (10)$$

which we use from here on.

2.1 Task setup and reward rate

We model a task in which the objective is to correctly identify which of two alternative stimuli have been presented on each trial so as to maximize the reward rate (RR) (Gold and Shadlen, 2002; Bogacz et al., 2006). We assume that all correct responses deliver equal reward, and that incorrect responses result in no reward. Therefore, the reward rate is the frequency with which correct responses are made, and is given by:

$$RR = \frac{1 - \langle ER \rangle}{\langle T \rangle} \quad (11)$$

where angle brackets denote averages over trials. ER is the error rate on the decision task, so that the numerator is the average fraction of correct responses in a run of trials. T is the response time of the model, measured starting from the beginning of the trial (before the stimulus onset at the random time t_d). Hence, RR is the expected fraction of correct responses divided by the average time elapsed for each response, and has units of 1/time.

In defining RR as above, we must take into account (1) how ‘premature’ responses made in the period $[0, t_d]$ (that is, in the RSI between the previous response and the presentation of the current stimulus) are scored as correct vs. incorrect, and (2) what effect, if any, these responses will have on the delay before presentation of the next stimulus. In deriving Eqn. (11), we adopt the following protocol, based on task designs in common use (including in (Clayton et al., 2004)): all premature responses are counted as errors, and the next trials are started immediately.

Stimulus presentations involve a step in stimulus strength a entering the first layer from zero to $a = +\bar{a}$, corresponding to stimulus 1 ($a_1 > a_2$), or to $a = -\bar{a}$ ($a_2 > a_1$), corresponding to stimulus 2. These stepped changes occur at randomized times t_d in each trial, and no other changes in stimulus strength occur. We take t_d to be uniformly distributed between 1 and 3 seconds in accord with empirical studies (Clayton et al., 2004).

Simulations were also performed with other ranges of the stimulus presentation time t_d (including fixing t_d to take only a single value, nonetheless assumed unknown to the decision maker) and other protocols for treating premature responses (such as simply ignoring any threshold crossing prior to stimulus presentation time t_d , at which point responses are immediately recorded if the firing rate difference $|y|$ remains above threshold). These variants had little qualitative impact on the results we report below comparing the fixed vs. dynamic gain models.

The standard parameter set

All parameters other than the ones of interest (i.e., gain and threshold values) are held fixed as RR is optimized for the various models. We refer to these values as the standard parameter set. First, we set the time constant $\tau = 1$ *sec.* and the weight $w = \beta$, as discussed above. We also take the signal-to-noise ratio $\left(\frac{\bar{a}}{\sqrt{\tau c}}\right)^2 = 8$ *sec.*⁻¹, a value derived from maximum likelihood fits of reaction time distributions from two-alternative choice task experiments in (Bogacz et al., 2006) (in the simulations, we use values $\bar{a} = 2$ and $c = 1/\sqrt{2}$, although any combination of \bar{a} and c in the same proportion would give the same optimized RR values; see Appendix). Finally, we take delays t_d between behavioral response and presentation of the next stimulus (response stimulus interval; RSI) to be uniformly distributed between 1 and 3 *sec.*

3 Optimization procedure

3.1 Optimizing decision models with fixed vs. dynamic gain mechanisms

Single layer model

For the one layer dynamic gain model, the free parameters are the gain values g_y^{pre} and g_y^{post} and the thresholds h_g and h . Thus, RR is a function of these parameters, via Eqn. (4), and its maximum value is given by

$$RR_{dynamic}^{1\ layer} = \max RR(g_y^{pre}, g_y^{post}, h_g, h) \text{ under Eqn. (4) .} \quad (12)$$

Next, we compare optimal RR s for the fixed and dynamic gain models by maximizing RR under fixed gain, and comparing this with the maximum RR obtained with a dynamic gain mechanism, as described above. The free parameters that remain are the gain value g_y^{const} and the response threshold h , and we find:

$$RR_{const}^{1\ layer} = \max RR(g_y^{const}, h) \text{ under Eqn. (4) .} \quad (13)$$

Two-layer model

For the two-layer dynamic gain model, there are four free gain parameters ($g_y^{pre}, g_z^{pre}, g_y^{post}, g_z^{post}$) and two thresholds (h_g and h). The maximum reward rate is:

$$RR_{dynamic}^{2\ layer} = \max RR(g_y^{pre}, g_z^{pre}, g_y^{post}, g_z^{post}, h, h_g) \text{ under Eqns. (9-10) ,} \quad (14)$$

In the two layer dynamic gain model, the gain transient affects both layers. In the brain, the extent to which norepinephrine adjusts the gain of a particular cortical population depends on the density and type of norepinephrine receptors in that population (as well as, e.g., connectivity within and afferent inputs to that population) (Waterhouse, Moises, and Woodward, 1998). In our model, however, the simplifying assumption is that the phasic LC impulse results in gain changes of the same magnitude in both the first and second layers. That is, a delay t_{NE} following the time T_g at a gain threshold is crossed in the first layer, $g_y^{pre} \rightarrow g_y^{pre} + \Delta g$ and $g_z^{pre} \rightarrow g_z^{pre} + \Delta g$. This reduces by one the number of free parameters in the optimization problem, but does not affect the generality of the results with regard to the eliminated parameters discussed in the Appendix. The maximum possible reward rate for the dynamic gain model becomes:

$$RR_{dynamic}^{2\ layer} = \max RR(g_y^{pre}, g_z^{pre}, \Delta g, h, h_g) \text{ under Eqns.(9)-(10) .} \quad (15)$$

We found that the maximal RR values found for Eqn. (15) and the slightly more general Eqn. (14) give the same optimal RR 's within $\sim 2\%$, so we adopt the simpler case of Eqn. (15) below.

For the two layer fixed gain model, there are three free parameters, g_y^{const} , g_z^{const} , and the response threshold h . The maximal RR found as these parameters are optimized is thus defined as:

$$RR_{const}^{2\ layer} = \max RR(g_y^{const}, g_z^{const}, h, h_g) \text{ under Eqns. (9)-(10) .} \quad (16)$$

3.2 Numerical optimization

We employed the SUBPLEX optimization procedure (Rowan, 1990) to optimize the free parameters in each model, thereby determining the maximal achievable reward rates for each. SUBPLEX is a variant of the Nedler and Mead (1965) simplex optimization procedure adapted for noisy systems. Specifically, RR (Eqn. (11)) was evaluated for each set of parameters sampled by the algorithm by Monte-Carlo simulation of 200,000 ‘trials’ of the appropriate model. The standard deviation in RR for a typical evaluation with this number of trials is 0.0003 sec^{-1} . The optimization procedure was seeded with different sets of randomly selected starting parameter values for each of the four model cases.

3.3 Range of reward rate values

We compared the maximal RR achieved by the optimization procedure for each model to theoretical ‘‘floor’’ and ‘‘ceiling’’ (putatively optimal) reward rates achievable in the two-alternative choice task described by our stimulus timing parameters and signal-to-noise ratio. First, recall that when stimulus onset time (t_d in our models) is known, the sequential probability ratio test (SPRT) provides the optimal performance for a two-alternative choice task such as the one studied here (Bogacz et al., 2006). Using the analytic formula in (Bogacz et al., 2006), we compute the reward rate achieved by the SPRT as $RR_{SPRT} = 0.440 \text{ sec.}^{-1}$.

As we described in the Introduction, the SPRT cannot be implemented for the case in which stimulus onset time is unknown, as is the case here. We are not aware of a statistical test that is known to be optimal in this setting. However, the CUSUM algorithm has been shown to be optimal under asymptotic limits of low error rates (i.e., large thresholds), and compares favorably with other statistical tests under more general conditions (see Chapter 5 of (Basseville and Nikiforov, 1993) and references therein). The ‘‘two-sided’’ version of this algorithm (Eqn. 2.2.24 of (Basseville and Nikiforov, 1993)) has separate integrators for the detection of two different possible changes in input

(corresponding to the two stimulus alternatives in our task), and we adopt this algorithm (using symmetric thresholds for each of the two integrators) as the primary benchmark against which we evaluate our decision models. With the stimulus detection threshold numerically optimized to maximize reward rate for the stimulus timing and signal-to-noise-ratios used in our simulations, the CUSUM algorithm gives $RR_{CUSUM} = 0.344 \text{ sec}^{-1}$. We take RR_{CUSUM} to be the ‘ceiling’ value RR_{ceil} of the reward rate achievable in the task we model.

To calculate the lower bound RR_{floor} of reward rate achievable in our task, we consider the RR achieved by chance guessing. We assume a decision maker who randomly chooses alternative 1 or 2 without regard to actual stimulus identity or onset, at a time uniformly chosen from 0 to 3 seconds (the maximal time of stimulus onset for our task). Since stimuli never appear prior to one second post trial onset, a third of these responses will be anticipatory by definition and thus regarded as errors. The remaining two-thirds of responses will fall in the 1 to 3 second interval, but will also be anticipatory 50% of the time, and thus erroneous. The remaining responses will be elicited post-stimulus, and will be correct 50% of the time. Thus, our rate of correct responding is given by $1 - ER = \frac{1}{3} * 0 + \frac{2}{3} * \frac{1}{2} * \frac{1}{2} = \frac{1}{6}$. The mean elapsed time between responses using this strategy is $\frac{3}{2}$ sec. Dividing these quantities as per Eqn. (11) yields $RR_{floor} \approx 0.111 \text{ sec}^{-1}$.

Given these observations, we define the range of reward rate achievable in our task as $[RR_{floor}, RR_{ceil}] = [.111, .344] \text{ sec}^{-1}$ for the standard parameter set. Below, we consider what proportion of this dynamic range of reward rate is achieved by each of our models, providing a relative measure of each model’s approach to optimality.

We note that our derivation of these bounds is intended to provide an intuitive framing of the reward rates that our models achieve relative to some theoretically motivated limits. In particular, our calculation of a lower bound for achievable reward rate assumes an agent that is motivated to perform, but agnostic about stimulus onsets. Naturally, an agent with an arbitrarily high threshold (i.e., not engaged in the task at all) will achieve a reward rate of zero by simply never responding. Our choices are motivated by framing our models’ performance within bounds that put in context the impact of either a fixed or dynamic gain mechanism on task performance.

4 Main Results

Here we report our results from optimizing model parameters to achieve maximal reward rate for the one- and two-layer models, with the fixed vs. dynamic gain mechanisms applied to each.

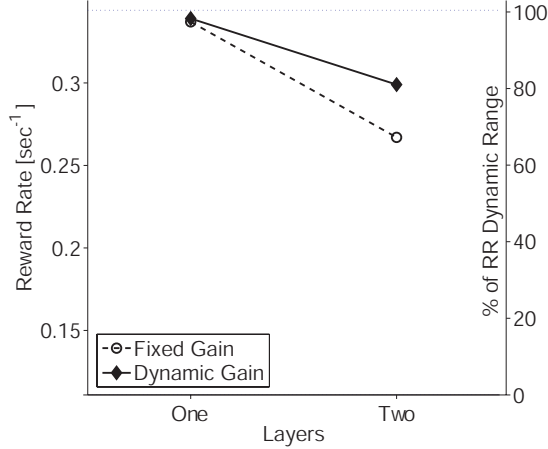


Figure 4: Summary of results for reward rate (RR) optimization. For the two-layer model, the optimization problem with dynamic gain gives higher reward rate than its fixed gain counterpart of Eqn. (16). However, the optimization problems for the single layer model yield similar reward rates for both the fixed gain and dynamic gain implementations. Reward rates are presented as percentages within the RR ‘dynamic range’ $[RR_{floor}, RR_{ceil}]$ (Sect. 3.3) along the right side axis. All results are for the standard parameter set of Sect. 2.1.

4.1 Optimal reward rates

Our main results are the maximal reward rates attained when each model’s parameters were independently optimized. These are presented graphically in Figure 4, which shows the maximal RR achieved by each model, along with the proportion of range of reward (Sect. 3.3) that the model was able to capture. The one layer fixed gain model achieved a maximal reward rate of $RR_{const}^{1\ layer} = 0.337\ sec.^{-1}$, whereas the one layer dynamic gain model reached a maximum of $RR_{dynamic}^{1\ layer} = 0.339\ sec.^{-1}$: a negligible difference. Therefore our model of LC-mediated dynamic gain has no meaningful effect on performance for a one layer decision mechanism. Nevertheless, note that both of these RR values are 97-98% of the RR range, indicating that a single layer model (with or without dynamic gain) closely approximates the performance of the CUSUM algorithm.

The situation for the two-layer model is quite different. For fixed gain, this model achieved a maximal value of $RR_{const}^{2\ layer} = 0.267\ sec.^{-1}$, equivalent to 67.0% of the theoretical range of reward rate. With dynamic gain, however, the two-layer model achieved $RR_{dynamic}^{2\ layer} = 0.299\ sec.^{-1}$, equivalent to 80.7% of this range.

To summarize, the fixed and dynamic gain implementations of the single

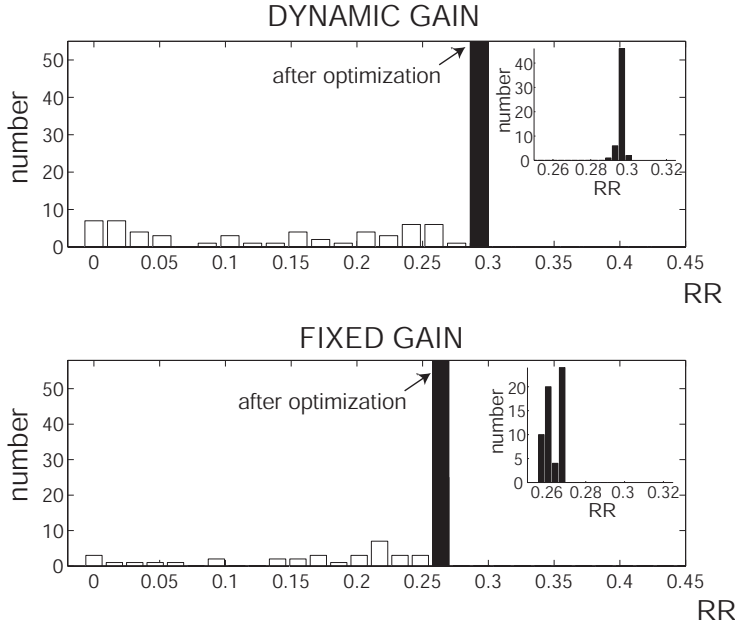
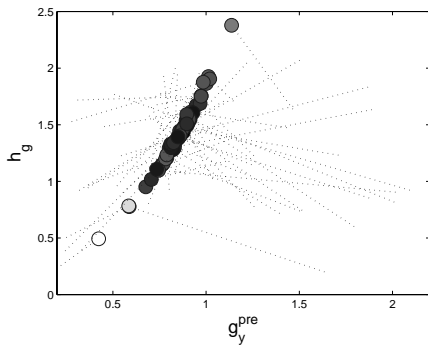


Figure 5: White bars: histogram of reward rates for randomly chosen, non-optimized parameter values. Filled bars: histogram of reward rates found using the SUBPLEX optimization algorithm on different runs, each starting from one of the non-optimized parameter sets. Top: Results for the two layer, dynamic gain optimization problem of Eqn. (15). Bottom: For the two layer, fixed gain optimization problem of Eqn. (16). Insets give zoomed view of optimized RR values with smaller histogram bin size (to illustrate the consistency of optimization results). Maximal reward rate obtained for the dynamic gain case is $RR_d=0.299 \text{ sec.}^{-1}$ (rightmost filled bar, top insert); maximum for fixed gain is $RR_c=0.267 \text{ sec.}^{-1}$ (rightmost filled bar, bottom insert).

layer model attained equivalent reward rates when optimized, and this value was higher than either of the two-layer model implementations. By contrast, using dynamic gain in the two layer system recovered an additional $\sim 14\%$ of the available reward, relative to the fixed gain model. We discuss implications of these results below, but first examine them in greater detail.

Figure 5 shows the distribution of reward rates achieved by the two layer, fixed and dynamic gain models before and after optimization. From the insets in this figure, it can be seen that the SUBPLEX algorithm converged to slightly different RR values each time it was run (for the two layer dynamic gain model, but not for the fixed gain model, there was a small minority of runs for which the algorithm did not converge; these runs are excluded). Nevertheless, this figure illustrates that the optimization algorithm generally arrived at consistent

A.



B.

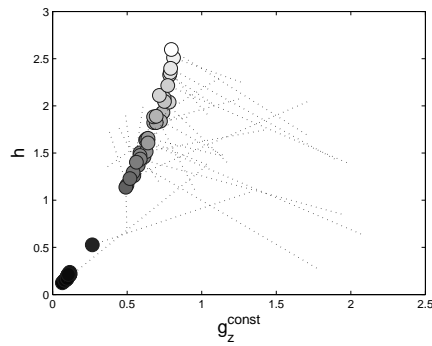


Figure 6: Optimization of parameter values for the two-layer model. Initial values of the parameters for each randomly initialized run of the algorithm lie at one end of the dotted lines, with final values indicated by the large dot at the other end. Color code of dots indicates obtained value of RR , with darker shades for higher values. (A) For the dynamic gain model, with RR values between 0.286 (white) and 0.299 (black) sec^{-1} . (B) For the fixed gain model, with RR values between 0.256 and 0.267 sec^{-1} .

optimized reward rates, regardless of starting parameter set.

Some of the optimized parameters leading to these RR values are shown in Figure 6(A) for the two-layer, dynamic gain model. Note that there is a family of different parameters that give similar RR values. Within this family, values of the *first-layer* gain threshold h_g increase with the gain g_y^{pre} , almost lying along a single curve. This may be understood as preserving the extent of evidence in favor of one or the other hypothesis necessary to trigger a LC-induced gain transient, as values of $|y|$ in Eqn. (10) will rise to greater values for larger g_y^{pre} . Fig. 6(B) gives analogous results for the fixed gain case, showing that optimized *second-layer* response thresholds h increase with g_z^{const} , falling along a similar curve. This suggests that, at optimal performance, gain thresholds in the first layer of dynamic gain models play a similar role to response thresholds in the second layer of fixed gain models. This property will be further illustrated in Fig. 7(A) below.

4.2 Behavioral predictions

We now describe the behavioral statistics that characterize the two layer model when it is optimized to produce optimal RR values via dynamic gain. We use the parameter set giving the best RR among all of those to which the SUBPLEX algorithm converged (cf. Fig. 6B): $g_y^{pre} = 0.873$, $g_z^{pre} = 0.474$, $\Delta g = 3.33$, $h_g = 1.43$, and $h = 1.86$.

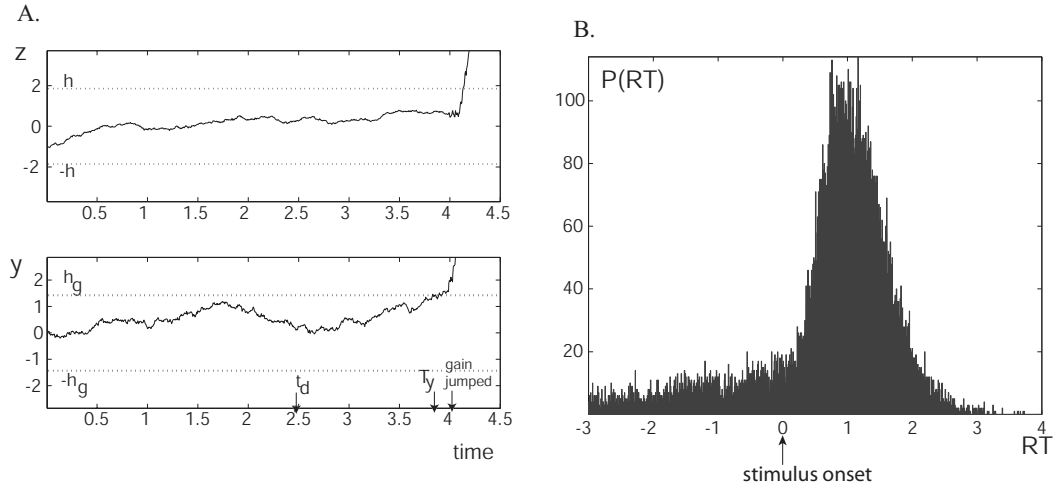


Figure 7: A. Typical trajectories at the decision (y) and response (z) layers in the optimized two-layer decision model (9-10) with dynamic gain. B. Corresponding distribution of reaction times, relative to stimulus presentation; negative values correspond to premature responses.

Typical trajectories for the firing rate equations (9-10) with these optimal parameters are shown in Fig. 7A. Note that, due to the relatively large value of Δg , Eqns. (9-10) become strongly unstable following the jump in gain (i.e., a delay t_{NE} after the thresholds $\pm h_g$ are crossed), and the response variable z rapidly crosses threshold. Figure 7B displays the response time distribution (stimulus-locked times $T - t_d$) resulting from an ensemble of such trials. Note that 2.0% of these were errors, while 16.8% were premature responses (i.e., occurring prior to stimulus onset). This rate of premature responses represents an extreme case, since the model assumed no information about the time of stimulus onset. In practice, subjects often have some information about stimulus timing, and/or can rapidly detect this from transients in stimulus energy (irrespective of identity). We will return to this issue further on.

Finally, Figure 8B shows that the times T_g at which the thresholds $\pm h_g$ were crossed are more tightly correlated with behavioral response than with stimulus onset times. This finding corresponds with the empirical observations made in (Clayton et al., 2004) that LC phasic responses correlate more closely with the time of behavioral response than stimulus onset for monkeys performing a two-alternative forced choice task (Clayton et al., 2004) (Fig. 8A). However, we emphasize that in the present study, no knowledge of the time at which stimulus salience develops is assumed, while such information could presumably be implicitly derived in the experiments of (Clayton et al., 2004) from the sudden onset of the task cues themselves. Thus, the present model

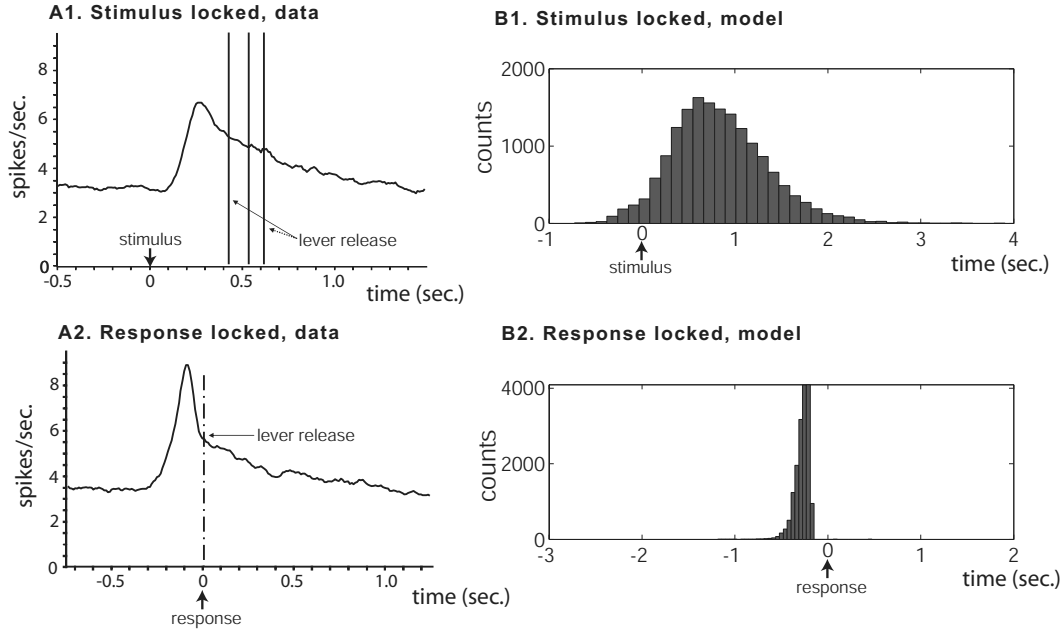


Figure 8: A. Peri-event time histograms (PETHs) of LC firing rates compiled across many experimental trials of the Eriksen flanker sensory discrimination task. Note that the observed post-stimulus transient increase in LC discharge is more tightly correlated with behavioral responses than stimulus presentations. Data kindly provided by E. Clayton and co-authors. B. Distribution of LC phase response latencies for the optimized two layer model with dynamic gain. Histograms show times T_g when the LC fires simulated phasic burst. B1 shows latencies relative to stimulus onset time t_a , and B2 relative to response time T .

should not be viewed as directly addressing the specific conditions of their task. Rather, it should be viewed as examining a limiting condition under which transients in gain mediated by the LC phasic response can impact performance. Finally, we note that response locking of LC responses has also been also seen in rodent odor conditioning tasks (Bouret and Sara, 2004; Bouret and Sara, 2005).

5 Discussion

The findings we have reported indicate that dynamic gain modulation, thought to be mediated by NE release from the locus coeruleus (LC), can help optimize performance on simulated sensory discrimination tasks – even when no knowledge of stimulus timing is assumed. The first model we considered had a single layer that both integrated differential evidence supporting one of two stimu-

lus identities, and elicited a decision when a response threshold was crossed. Gain transients served to accelerate trajectories toward the response threshold when sufficient evidence had accumulated (i.e., a gain threshold was crossed). We observed that optimizing the parameters of this model did not result in improved performance relative to a similar single layer model with fixed gain (Figure 4). Presumably, this reflects the fact that in a single layer model the same units are responsible for integrating stimulus information and signaling a response. Thus, it appears that once sufficient evidence has accumulated to detect the presence of a stimulus (and hence trigger a transient increase in gain), there is also sufficient evidence to identify the stimulus and signal a response. It is interesting to note in this context that changes in gain in a single layer model are closely related to, and under certain conditions can be equivalent to, changes in response threshold (Servan-Schreiber, Printz, and Cohen, 1990). Thus, our finding that allowing gain to vary failed to afford any improvements in performance suggests that the same may be true for varying thresholds in single-layer models².

We next considered two layer models in which decision (stimulus discrimination) and response processes were segregated into separate layers. This was motivated by the recognition that, in the brain, the mechanisms responsible for the decision process (e.g., stimulus discrimination in our task) are rarely if ever co-localized with those responsible for response execution. Here again we compared a fixed gain version of the model with one in which the decision process could elicit a transient increase in gain in both the decision and response layers. In contrast to the single layer model, we found that allowing a dynamic adjustment in gain afforded an improvement in performance.

Specifically, we found that the best performance was achieved when the decision layer could trigger a transient increase in gain once the threshold in that layer had been reached. As we have noted earlier, a properly parameterized single layer model can approximate optimal decision algorithms when stimulus onset is not known, at least in limiting cases (i.e., the CUMSUM test; see Section 3.3). Thus, any additional processing in the response layer is likely superfluous. This was evident in our results, which showed that even when the parameters of the fixed gain version of two layer model were optimized, it performed substantially worse than the single layer model (see Fig. 4). However, this can be ameliorated by allowing the system to increase gain once the integration process in the decision layer has crossed threshold. This has the effect of ‘binarizing’ units in the response layer, diminishing any further integration in that layer and forcing a response based on the current state of the decision layer. That is, the dynamic adjustment of gain effectively collapses the two layer mechanism to a single layer mechanism once a decision has been reached,

²We would like to thank Angela Yu for calling this point to our attention.

thereby recovering some of the efficiency of a single layer model. Our findings validate this claim, and show that it generalizes over a variety of modeling assumptions (as described in Appendix 1). Along these lines, it is worth noting that the threshold in the first layer of the optimally parameterized two layer model (with dynamic gain) is close to the one in the optimally parameterized single layer model (with fixed gain).

When gain is held constant in our model, processing occurs on roughly similar timescales in both of its layers. This reflects an assumption, discussed in more detail below, that both model layers are capable of accumulating their inputs over time (although they are not necessarily pure integrators, *i.e.* implementing the DDM, depending on their gain levels). However, alternative two-layer models are also possible in which the second layer operates much differently, processing its inputs on much faster timescales than the first and perhaps with little additional noise. In such cases, there will be less superfluous processing in the second layer for the LC to ameliorate, and the present dynamic gain mechanism will likely have substantially lesser effect.

Relationship to previous studies

Our results are consistent with the hypothesis that phasic LC responses may mediate dynamic adjustments of gain. Previous studies have demonstrated that the effects of NE release at both the physiological and behavioral levels can be modeled as a modulation of gain (e.g., (Servan-Schreiber, Printz, and Cohen, 1990; Usher et al., 1999)). Here we have shown that when gain transients are optimized to increase reward rate in the absence of stimulus timing information, the timing of such transients is more tightly correlated with the timing of behavioral response than stimulus onset. This agrees with empirical observations concerning the response locking of LC phasic responses, as well as their relationship to behavioral performance (Clayton et al., 2004; Bouret and Sara, 2004; Bouret and Sara, 2005) (although we emphasize again that the stimulus condition modeled here cannot be directly compared with that of (Clayton et al., 2004), in which subjects could presumably infer stimulus timing from strong onset effects of the cue itself).

Additionally, our results are related to the work of (Dayan and Yu, 2006). These authors interpret phasic impulses of norepinephrine as representing the probability that a target stimulus is present. Indeed, in our model gain (and hence LC) transients are elicited only when sufficient evidence has been accumulated determine that a specific task alternative is present, and hence may be viewed as a discrete analog of the graded LC responses in (Dayan and Yu, 2006). The change in processing that follows LC transients in our model is also related to the general concept of LC-driven network “interrupts” or “resets” in (Dayan and Yu, 2006) and (Bouret and Sara, 2005).

Elsewhere, we have proposed that the adaptive regulation of gain by the LC may serve as a temporal filter, facilitating responses to task-relevant information at the moment that such information is being actively represented (Aston-Jones and Cohen, 2005) and thereby improving performance. That said, the improvements in performance that we observed for the dynamic gain mechanism were modest, and did not fully reach the near optimal levels achieved by single layer models (Figure 4). This might call into question the value of such a mechanism, relative to its implementational “expense.” However, this must be considered in the context of several other observations and factors, which we now undertake.

Interpretation of optimized reward rates

First, as shown in Appendix 1, the dynamic gain mechanism affords substantially greater improvements in performance when response thresholds are fixed at a high level. Under these conditions, a two layer model with dynamic gain adjustment can achieve a reward rate comparable to that of its single layer counterpart. At present, it is not known how response thresholds are implemented in real neural architectures (see (Simen, Cohen, and Holmes, 2006; Lo and Wang, 2006) for recently proposed models), and whether they are adapted to specific tasks, or remained fixed at a high level. Our present findings suggest that, in either case, dynamic gain adjustment provides at least modest benefits, that become more valuable to the extent that response thresholds assume high, “protective” values.

A second consideration is the overall complexity of the processing network. We have considered the two simplest cases: single and two layer models. The latter can be considered the simplest, and most conservative proxy for the effects of dynamics gain adjustment in more complex networks. Such multilayer networks may arise because different tasks require integration and/or discrimination of information of different types, from different sources, and at different levels of analysis. For example, one task may require a decision based on (and therefore accumulation of) information about direction of motion. Another might require accumulation of information about object identity, regardless of motion. And a third might require combining different sources of information or levels of analysis (e.g., separately identifying direction of motion and target location in the experiment of (Sato et al., 2001)). Thus, although the decision process for a given task may be optimally implemented as a single layer network, the mechanisms responsible for actual performance of a given task are likely implemented by different parts (“layers”) of the system that may be imbedded at different levels within the full neural architecture (Sato et al., 2001; Reddi, 2001; Schall, 2003), varying in the degree to which they have direct access to response mechanisms.

It seems reasonable to assume that, if additional layers of processing are interposed between the decision and response layers, unnecessary integration at each of these intermediate layers will further degrade performance. Thus, dynamic gain adjustment could become increasingly important as the complexity of the processing system increases. However, it is worth noting that the LC mechanism itself does not need to increase in complexity to meet such demands, given its ability to effect rapid, widespread gain modulation. The only demand is that the system be able to identify the decision layer relevant to a given task, that this be able to optimally adapt its decision threshold, and that threshold crossing in this layer drive the LC. The first two of these are requirements for task performance, irrespective of the involvement of LC (and are presumably subserved by mechanisms responsible for cognitive control - cf. (McClure, Gilzenrat, and Cohen, 2005b; Simen, Cohen, and Holmes, 2006)). The last requirement - that the decision layer in a given task be able to drive LC phasic responses - is an interesting and important consideration. There are at least two ways in which this could occur. One is through the effects of top-down control, and the other is learning. As just noted, control mechanisms are required in any event to identify relevant processing pathways for task execution, and thus may also serve to direct the output of task-relevant decision layers to the LC. Such pathways may also be subject to learning. The latter is the focus of recent modeling and empirical work (McClure, Gilzenrat, and Cohen, 2005a).

Finally, even if we allow that dynamic gain adjustment provides only modest increases in reward rate, the impact of such benefits may be important in highly competitive environments (G. Aston-Jones, personal communication). In such environments (e.g., when resources are scarce) even small advantages, if they are consistent, can be amplified and reinforced under selective evolutionary pressure. This is particularly true if the necessary mechanisms can emerge from simple modifications of existing, evolutionarily conserved mechanisms that serve other functions. This is almost certainly the case for the LC-NE system, which is a phylogenetically old neural system. Elsewhere, we have proposed that LC phasic responses may represent one function of a system that is adapted more generally to regulate the balance between exploitation and exploration - a fundamental challenge that confronts all behaving organisms, from fungi to ant colonies (Watkinson et al., 2005; Pratt and Sumpter, 2006; Cohen, McClure, and Yu, 2007).

Conclusions

Identifying and responding to noisy stimuli requires the relatively slow accumulation of evidence to determine an outcome. However, once this outcome is reached, additional layers of processing between decision and response areas

may serve primarily to route the outcome to the appropriate behavior. We have shown that an appropriately timed global gain transient can streamline this process by sensitizing the pathway at the right time. This transient potentiation of upstream processing areas serves to “collapse” the additional layers of processing around the current state of the decision layer, effectively recovering some of the performance advantages of an optimal, single layer architecture. Our findings provide a more precise formal understanding of prior simulation results suggesting that LC-NE system implements such a dynamic gain adjustment mechanism, and lend further support to the idea that this system may play an important role in optimizing decision making in the brain.

Acknowledgements. This work was partially funded by the grant NIH P50 MH62196 (Cognitive and Neural Mechanisms of Conflict and Control, Silvio M. Conte Center). E.B. and M.G. were supported under National Science Foundation Graduate Fellowships and E.B. under a Burroughs-Wellcome Fund (BWF) Training Grant in Biological Dynamics and BWF Career Award at the Scientific Interface. The authors are grateful to Josh Gold and Rafal Bogacz for useful contributions and discussions, as well as to Ed Clayton and Gary Aston-Jones for providing the data of Figure 8 and for their insights into the role of the LC in modulating decisions. The authors thank Phil Holmes for comments on the manuscript, and the reviewers for suggestions that have substantially improved our paper.

Appendix 1: Extensions to other constraints and models

Here, we consider modifications to the decision model that affect the implementation of noise, the range over which parameters may vary, and the manner in which neural gain enters the governing equations. These leave the main results mostly unchanged.

Fixed response thresholds

Here, we examine an alternative model with a fixed response threshold. This model implements the more conservative assumption that the response threshold h is not adjusted from task to task, but rather remains fixed at some sufficiently high value as to protect against premature crossing over a variety of task conditions (i.e., varying signal-to-noise ratios and stimulus magnitudes³). To implement this assumption, we fix $h = 5$, roughly 2-3 times typical threshold values for near-optimal solutions for the case of Eqn. (15) in which h was allowed to vary freely.

For one layer models with fixed response threshold $h = 5$, the optimal performance for the dynamic gain model is $RR=0.306 \text{ sec.}^{-1}$, whereas for the fixed gain model it is $RR=0.281 \text{ sec.}^{-1}$. These values correspond to 83.7% and 73.0% of the dynamic range $[RR_{floor}, RR_{ceil}]$, respectively. Thus, when the response threshold remains fixed, dynamic gain yields a performance advantage for the one layer model. This contrasts with the situation for one layer models with freely varying response thresholds, in which dynamic gain did not improve performance.

For two layer models with fixed response threshold $h = 5$, the highest RR found with dynamic gain is $RR=0.299 \text{ sec.}^{-1}$. A maximal value of $RR=0.247 \text{ sec.}^{-1}$ is achieved for the fixed gain implementation; these values correspond to 58.4% and 80.7% of the dynamic range. Thus, dynamic gain modulation allows expected rewards to be enhanced by 22% over any fixed task interval. Furthermore, note that the RR value 0.299 sec.^{-1} obtained here with fixed response threshold and dynamic gain matches that obtained above for the case of Eqn. (15), which allowed h to vary freely. Thus, the two layer model equipped with a dynamic gain mechanism can maintain similarly high levels of performance irrespective of whether response thresholds are fixed or not, while fixing the response thresholds compromises performance when gain is constant.

³We thank Josh Gold for this suggestion.

Stimulus-dependent noise

Thus far, we have considered only models in which noise levels remain constant before and after stimulus presentation, attaining the signal-to-noise ratio $\sim \bar{a}/c$ (given by our standard parameter set) at time t_d . Here, we consider the effectiveness of dynamic gain mechanisms in a case where noise varies with stimulus strength. Specifically, we set $c = \frac{1}{4} * \frac{1}{\sqrt{2}}$ for $t < t_d$ and $c = \frac{1}{\sqrt{2}}$ for $t \geq t_d$. The factor of $\frac{1}{4}$ is an arbitrary value chosen to illustrate a representative example.

The optimal reward rate obtained for the two layer model with dynamic gain is $RR=0.361 \text{ sec.}^{-1}$. For comparison, a maximum of $RR=0.311 \text{ sec.}^{-1}$ was obtained for the analogous two-layer model with fixed gain. Following numerical optimization with stimulus-dependent noise, we find $RR_{CUSUM} = 0.435 \text{ sec.}^{-1}$. Therefore, using the same lower bound as before, dynamic gain yields an improvement of 15% in the dynamic range of available reward. We conclude that the dynamic gain mechanism continues to afford significant improvements in reward rate in the presence of stimulus-dependent noise.

Another model for neural integrators

Here we consider the robustness of our findings to assumptions about the neural integration process. Specifically, we consider the following leaky integrator model of the decision task (McClelland, 1979; Usher and McClelland, 2001):

layer 1:

$$\tau \frac{dx_1}{dt} = -x_1 - \beta f_{g(t)}(x_2) + a_1(t) + \frac{c(t)}{\sqrt{2}} \sqrt{\tau} \eta_1(t) \quad (17)$$

$$\tau \frac{dx_2}{dt} = -x_2 - \beta f_{g(t)}(x_1) + a_2(t) + \frac{c(t)}{\sqrt{2}} \sqrt{\tau} \eta_2(t) \quad (18)$$

layer 2:

$$\tau \frac{dv_1}{dt} = -v_1 - \beta f_{g(t)}(v_2) + w f_{g(t)}(x_1) + \frac{c(t)}{\sqrt{2}} \sqrt{\tau} \eta_3(t) \quad (19)$$

$$\tau \frac{dv_2}{dt} = -v_2 - \beta f_{g(t)}(v_1) + w f_{g(t)}(x_2) + \frac{c(t)}{\sqrt{2}} \sqrt{\tau} \eta_4(t) \quad (20)$$

where the state variables $x_j(t)$ and $v_j(t)$ denote the mean input currents to cell bodies of the j th neural unit in the decision and response layers, respectively. The integration implicit in the differential equations represents temporal summation of dendritic synaptic inputs ((Grossberg, 1988) and references therein). Unit firing rates are given by $f_{g(t)}(x_j(t))$, where $f_{g(t)}(\cdot)$. The other terms are as for Eqns.(1-2) and (5-6). The main difference between the present model ((17-20) and the model ((1-2), (5-6)) studied in earlier sections is whether all

of the deterministic inputs, or just the activity of the opposing unit, appear inside the function $f_{g(t)}(\cdot)$. See (Brown et al., 2005), cf. (Grossberg, 1988), for more on the relationship between these models.

Equations for the differences $x = x_1 - x_2$ and $v = v_1 - v_2$ may be derived, as in Section 2, via linearization of the activation functions $f_{g(t)}(\cdot)$. After eliminating β and w exactly as in the Appendix, these equations are

$$\tau \frac{dx}{dt} = -x + g_x x + a + c\sqrt{\tau}\eta_2(t) \quad (21)$$

$$\tau \frac{dv}{dt} = -v + g_v v + g_x x + c\sqrt{\tau}\eta_1(t) \quad (22)$$

As above, we study only differences in firing rates. Following linearization in the second layer, $f_{g(t)}(v_1) - f_{g(t)}(v_2) = g(t)v$, and it is this quantity that we require to exceed one of the thresholds $\pm h$ for a response to occur (and similarly for $g(t)x$ and the triggering of LC response and concomitant gain increases following crossing of thresholds $\pm h_g$). Thus, gain transients have two effects: (i) with respect to the variables (x, v) , the effective thresholds are adjusted and (ii) the linear terms in the equations (21-22) are themselves changed.

In order to assess the extent to which dynamic gain strategies can improve performance in the reduced model (21-22) and its single layer analog, we define reward rate optimization problems corresponding to those of Section 3. That is, for the one layer dynamic gain model, we optimize RR over free parameters g_x^{pre} and g_x^{post} and the thresholds h_g and h . For the corresponding one layer model with fixed gain, we optimize RR over free parameters g_x^{const} and the response threshold h . For the two-layer dynamic gain model, we optimize RR produced by (21-22) over the gain parameters $(g_x^{pre}, g_v^{pre}, \Delta g)$ and over the thresholds $(h_g$ and $h)$. Finally, for the two layer fixed gain model, RR is optimized over g_x^{const}, g_v^{const} , and h .

The results are as follows. First, note that, using coordinate transformations and rescalings similar to those in the Appendix, it may be shown that the fixed gain reward rate optimization problem for the leaky accumulator model (21-22) is actually equivalent to its analog for the firing rate model (7-8). Therefore, it is not surprising that numerical optimization with the SUBPLEX algorithm yields $RR = 0.337 \text{ sec.}^{-1}$ for the single layer model and $RR = 0.267 \text{ sec.}^{-1}$ for the two-layer model, exactly as for the firing rate model in Section 4.

For the dynamic gain problem, the same value of $RR = 0.337 \text{ sec.}^{-1}$ is again obtained for the single layer model (21) as with fixed gain. Thus, we find once again that dynamic gain does not yield a performance advantage for the one layer decision model. For the two-layer model with dynamic gain, we obtain $RR=0.316 \text{ sec.}^{-1}$. This value of RR is slightly higher than the one

obtained for the firing rate model (9-10), a finding attributable to the twin effects of Δg mentioned above for the present model.

The main implication of the results described in this appendix is that the qualitative effects of the dynamic gain mechanism do not depend critically on the details of how gain enters into the model equations. Thus, even though we do not yet have a precise biophysical model of how NE release produces gain modulation, our findings suggest that this modulation will lead to improved performance under relatively general conditions.

Appendix 2: Parameters in decision model and generality of optimization results

Here, we discuss the role of parameters in the decision models. In particular, we identify two parameters (β and k) whose values are inconsequential to the optimization problems. That is, the models would produce the same optimal reward rates regardless of the value of these parameters, an observation that eliminates the need to specify their value and extends the generality of our results. We also discuss the choices we make for the parameter w , and how the models would change if another choice had been made.

We start with the two-layer model, Eqns. (7)-(8). First, we show that the scale factor k (which sets the absolute magnitude of stimulus and noise inputs) may be eliminated by rescaling thresholds h and h_g . Dividing Eqns. (7) and (8) through by k and defining $\hat{y} = y/k$, $\hat{z} = z/k$ we obtain

$$\tau \frac{d\hat{z}}{dt} = -\hat{z} + g_z \beta \hat{z} + w g_z \hat{y} + g_z c \sqrt{\tau} \eta^1(t) \quad (23)$$

$$\tau \frac{d\hat{y}}{dt} = -\hat{y} + g_y \beta \hat{y} + g_y a + g_y c \sqrt{\tau} \eta^2(t) . \quad (24)$$

After scaling thresholds h and h_g by the (positive) factor $1/k$ (matching the scaling of y), we obtain exactly the same threshold crossings for Eqns. (23)-(24) as for the original Eqns. (7)-(8); furthermore, since thresholds are free parameters in the optimization problems we consider, we have lost no generality in eliminating the parameter k .

We additionally note that dividing Eqns. (23)-(24) by c , and rescaling \hat{y} , \hat{z} (and the associated thresholds) by $1/c$ would demonstrate that only the signal to noise ratio $\sim a/c$ affects optimal reward rates, not the absolute values of these parameters.

Next, making the definitions $\tilde{g}_y = \beta g_y$, $\tilde{g}_z = \beta g_z$, $\tilde{y} = \beta \hat{y}$, and $\tilde{z} = \beta \hat{z}$, and

multiplying both equations by β , we get

$$\tau \frac{d\tilde{z}}{dt} = -\tilde{z} + \tilde{g}_z \tilde{z} + \frac{w}{\beta} \tilde{g}_z \tilde{y} + \tilde{g}_z c \sqrt{\tau} \eta^1(t) \quad (25)$$

$$\tau \frac{d\tilde{y}}{dt} = -\tilde{y} + \tilde{g}_y \tilde{y} + \tilde{g}_y a + \tilde{g}_y c \sqrt{\tau} \eta^2(t) . \quad (26)$$

Recall that the parameters \tilde{g}_y, \tilde{g}_z can take arbitrary (positive) values in the optimization schemes, so that rescaling by the constant $\tilde{\beta}$ has no effect on the optimization problem. Rescaling thresholds \tilde{h}, \tilde{h}_g again and applying the fact that they are free variables under optimization, we see that the system of Eqns. (25)-(26) also produces the identical optimal error rates as the original Eqns. (7)-(8).

Finally, we multiply Eqn. (25) by $\frac{\beta}{w}$ and rescale z again, setting $z' = \frac{\beta}{w} \tilde{z}$. This gives the equations

$$\tau \frac{dz'}{dt} = -z' + \tilde{g}_z z' + \tilde{g}_z \tilde{y} + \tilde{g}_z \frac{\beta}{w} c \sqrt{\tau} \eta^1(t) \quad (27)$$

$$\tau \frac{d\tilde{y}}{dt} = -\tilde{y} + \tilde{g}_y \tilde{y} + \tilde{g}_y a + \tilde{g}_y c \sqrt{\tau} \eta^2(t) . \quad (28)$$

which are again equivalent to the former equations when optimized.

Equation (27) clarifies the role of the ratio $\frac{\beta}{w}$, which determines the noise strength in the second layer relative to the first. We make the choice $w = \beta$ for parsimony, making the noise strengths in the two layers equal. This yields the equations (9-10).

Finally, we note that the same arguments above, if applied to the single layer model of Eqn. (3), yield the simplified model Eqn. (4).

References

- Abbott, L. (1991). Firing-rate models for neural populations. In Benhar, O., C. Bosio, P. Del Giudice, and E. Tabat, editors, *Neural Networks: from Biology to High-Energy Physics*, pp. 179–196. ETS Editrice, Pisa.
- Aston-Jones, G. and J. Cohen (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28: 40350.
- Aston-Jones, G., J. Rajkowski, P. Kubiak, and T. Alexinsky (1994). Locus coeruleus neurons in the monkey are selectively activated by attended stimuli in a vigilance task. *J. Neurosci.* 14: 4467–4480.
- Basseville, M. and I. Nikiforov (1993). *Detection of Abrupt Changes: Theory and Application*. Prentice Hall, New Jersey.

- Bogacz, R., E. Brown, J. Moehlis, P. Holmes, and J.D. Cohen (2006). The physics of optimal decision making: A formal analysis of models of performance in two alternative forced choice tasks. *Psych. Rev.* 113: 700–765.
- Bouret, S. and S.J. Sara (2004). Reward expectation, orientation of attention and locus coeruleus medial frontal cortex interplay during learning. *Eur. J. Neurosci.* 20: 791802.
- Bouret, S. and S.J. Sara (2005). Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. *Trends in Neurosciences.* 28: 574–582.
- Brown, E., J. Gao, P. Holmes, R. Bogacz, M. Gilzenrat, and J Cohen (2005). Simple networks that optimize decisions. *Int. J. Bifurcation and Chaos* 15: 803–826.
- Brown, E. and P. Holmes (2001). Modeling a simple choice task: stochastic dynamics of mutually inhibitory neural groups. *Stochastics and Dynamics* 1(2): 159–191.
- Clayton, E., J. Rajkowski, J.D. Cohen, and G. Aston-Jones (2004). Phasic activation of monkey locus ceruleus neurons by simple decisions in a forced-choice task. *J. Neuroscience* 24: 9914–9920.
- Cohen, J.D., K. Dunbar, and J.L. McClelland (1990). On the control of automatic processes: A parallel distributed processing model of the Stroop effect. *Psychological Review* 97(3): 332–361.
- Cohen, J.D., S.M. McClure, and A.J. Yu (2007). Should I stay or should I go? how the human brain manages the tradeoff between exploitation and exploration. *Under review* .
- Dayan, P. and A. Yu (2006). Phasic norepinephrine: A neural interrupt signal for unexpected events. *Network: Computation in Neural Systems* 17: 335–350.
- Gerstner, W. and W. Kistler (2002). *Spiking Neuron Models*. Cambridge University Press.
- Gilzenrat, M.S., B.D. Holmes, J. Rajkowski, G. Aston-Jones, and J.D. Cohen (2002). Simplified dynamics in a model of noradrenergic modulation of cognitive performance. *Neural Networks* 15: 647–663.
- Gold, J. and M. Shadlen (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.
- Hanes, D.P. and J.D. Schall (1996). Neural control of voluntary movement initiation. *Science* 274: 427–30.
- Heeger, D.J., G.M. Boynton, J.B. Demb, E. Seidemann, and W.T. Newsome (1999). Motion opponency in visual cortex. *Journal of Neuroscience* 19: 7162–7174.
- 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.
- Laming, D.R.J. (1968). *Information Theory of Choice-Reaction Times*. Academic Press, New York.
- Lehmann, E.L. (1959). *Testing Statistical Hypotheses*. John Wiley & Sons, New York.
- Lo, C.-C. and X.-J. Wang (2006). Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nat. Neurosci.* 9: 956–963.
- Machens, C., R. Romo, and C. Brody (2005). Flexible control of mutual inhibition: A neural model of two-interval discrimination. *Science* 307: 1121 – 1124.
- McClelland, J.L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review* 86: 287–330.
- McClure, S., M. Gilzenrat, and J. Cohen (2005a). An exploration-exploitation model based on norepinephrine and dopamine activity. *Advances in Neural Information Processing Systems* 18.
- McClure, S., M. Gilzenrat, and J. Cohen (2005b). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience* 24: 167–202.
- Pratt, S.C. and D.J.T. Sumpter (2006). A tunable algorithm for collective decision-making. *Proc. Natl. Acad. Science, USA* 103(43): 15906–15910.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psych. Rev.* 85: 59–108.
- Ratcliff, R., A. Thapar, and G. McKoon (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception and Psychophysics* 65: 523535.

- Ratcliff, R., T. Van Zandt, and G. McKoon (1999). Connectionist and diffusion models of reaction time. *Psych. Rev.* 106 (2): 261–300.
- Reddi, B. (2001). Decision making: two stages of neural judgement. *Curr. Biol.* 11(15): R603–R606.
- Rowan, T. (1990). Functional Stability Analysis of Numerical Algorithms. Ph.D. diss., University of Texas at Austin.
- Sato, T., A. Murthy, K.G. Thompson, and J.D. Schall (2001). Search efficiency but not response interference affects visual selection in frontal eye field. *Neuron* 30: 1–20.
- Schall, J. (2003). Neural correlates of decision processes: neural and mental chronometry. *Curr. Opin. Neurobiol.* 13: 182–186.
- Schall, J. D. and K. G. Thompson (1996). Motion perception: Seeing and deciding. *Proc. Natl. Acad. Sci. USA* 93: 628633.
- Schall, J. D. and K. G. Thompson (1999). Neural selection and control of visually guided eye movements. *Annu. Rev. Neurosci.* 22: 241259.
- Servan-Schreiber, D., H. Printz, and J.D. Cohen (1990). A network model of catecholamine effects: Gain, signal-to-noise ratio, and behavior. *Science* 249: 892–895.
- Shadlen, M.N. and W.T. Newsome (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiology* 86: 1916–1936.
- Simen, P., J.D. Cohen, and P. Holmes (2006). Adaptation of decision making parameters by continuous performance monitoring. *Neural Networks* 19: 1013–26.
- Simoncelli, E.P. and D.J. Heeger (1998). A model of neuronal responses in visual area mt. *Vision Research* 38: 743–761.
- Smith, P.L. and R. Ratcliff (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosci.* 27 (3): 161–168.
- Stone, M. (1960). Models for choice-reaction time. *Psychometrika* 25: 251–260.
- Usher, M., J.D. Cohen, D. Servan-Schreiber, J. Rajkowski, and G. Aston-Jones (1999). The role of locus coeruleus in the regulation of cognitive performance. *Science* 283: 549–554.

- Usher, M. and E.J. Davelaar (2002). Neuromodulation of decision and response selection. *Neural Networks* 15: 635–645.
- Usher, M. and J.L. McClelland (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psych. Rev.* 108: 550–592.
- Wald, A. (1947). *Sequential Analysis*. John Wiley & Sons, New York.
- Waterhouse, B., H. Moises, and D. Woodward (1998). Phasic activation of the locus coeruleus enhances responses of primary sensory cortical neurons to peripheral receptive field stimulation. *Brain Res.* 790: 33–44.
- Watkinson, S.C., L. Boddy, K. Burton, P.R. Darrah, D. Eastwood, M.D. Fricker, and M. Tlalka (2005). New approaches to investigating the function of mycelial networks. *Mycologist* 19: 11–17.
- Wilson, H. and J. Cowan (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* 12: 1–24.
- Wong, K. F. and X. J Wang (2006). A recurrent network mechanism of time integration in perceptual decisions. *J. Neurosci.* 26: 1314–1328.