Neural Networks 24 (2011) 417–426

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Hebbian learning in linear–nonlinear networks with tuning curves leads to near-optimal, multi-alternative decision making

Tyler McMillen [a,*], Patrick Simen [b], Sam Behseta [a]

[a] Department of Mathematics, California State University at Fullerton, Fullerton, CA 92834, United States

[b] Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, United States

ARTICLE INFO

Article history:
Received 1 March 2010
Received in revised form 12 September 2010
Accepted 20 January 2011

Keywords:
Synaptic weight learning
Leaky accumulator
Drift–diffusion model
Neural network
Multihypothesis sequential test
Sequential ratio test

ABSTRACT

Optimal performance and physically plausible mechanisms for achieving it have been completely characterized for a general class of two-alternative, free response decision making tasks, and data suggest that humans can implement the optimal procedure. The situation is more complicated when the number of alternatives is greater than two and subjects are free to respond at any time, partly due to the fact that there is no generally applicable statistical test for deciding optimally in such cases. However, here, too, analytical approximations to optimality that are physically and psychologically plausible have been analyzed. These analyses leave open questions that have begun to be addressed: (1) How are near-optimal model parameterizations learned from experience? (2) What if a continuum of decision alternatives exists? (3) How can neurons' broad tuning curves be incorporated into an optimal-performance theory? We present a possible answer to all of these questions in the form of an extremely simple, reward-modulated Hebbian learning rule by which a neural network learns to approximate the multihypothesis sequential probability ratio test.

1. Introduction

In this work, we examine the problem of maximizing earnings from a sequence of N-alternative decisions about the identity of noisy stimuli, with $N \geq 2$. Our goal is to parameterize a simple neural circuit model whose behavior approximates optimal performance in such tasks, while simultaneously accounting for the fundamental role of tuning curves in the neural representation of sensory stimuli. Throughout, we take 'optimal' to mean *reward maximizing*, and we assume that correct decisions earn rewards for the decider. We also assume that rewards are maximized while subject to constraints of biological plausibility and computational tractability, which we take to be satisfied by stochastic neural networks whose size grows linearly with the number of decisions to be made.

As we show, simple principles of neural computation are sufficient to approximate this form of optimality in a class of N-choice tasks involving response-terminated stimuli: that is, stimuli that provide information continuously until the time (the response time) at which task participants decide for themselves when to stop observing and make a response. This is somewhat surprising, given that a general decision policy that guarantees truly optimal performance cannot even be explicitly formulated for such tasks, as we discuss below (Dragalin, Tartakovsky, & Veeravalli, 1999).

In the simple decision making tasks we consider, we assume that participants earn rewards for correct responses (correct stimulus identifications), and earn less for errors (for simplicity, we assume that errors earn nothing). Rewards may be explicitly provided by the experimenter, or else we may assume that participants arbitrarily assign some measures of personal satisfaction to correct responses that are not determined by the experimenter. Each stimulus type has a fixed prior probability within a block of trials, and the average signal-to-noise ratio of each stimulus is fixed. The block duration, rather than the number of trials, is also held fixed, and the distribution of response-to-stimulus intervals (RSIs) that delay the onset of the next stimulus after a response is stationary. In this case, maximizing the rate of reward also maximizes the total reward.

Maximizing gains in this and a variety of similar tasks requires probabilistic inference. While the importance of a principled inference process is widely understood in psychology and neuroscience (Knill & Pouget, 2004), the complexity of optimal decision policies in tasks with response-terminated stimuli (also known as 'free response' or 'response time' tasks) and $N > 2$ choices appears to be less well appreciated. In addition, the relationship between neural representations and decision variables takes on a new level of complexity when $N > 2$. In the $N = 2$ case, tuning curve width is irrelevant if one makes the simplifying assumption that the decision making system consists

* Corresponding author. Tel.: +1 657 278 8208.
E-mail addresses: tmcmillen@fullerton.edu (T. McMillen), psimen@math.princeton.edu (P. Simen), sbehseta@fullerton.edu (S. Behseta).

0893-6080/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2011.01.005

of a neuron tuned for one type of stimulus and an 'anti-neuron' tuned with the same sensitivity to the opposite stimulus (Gold & Shadlen, 2001). In contrast, when $N > 2$, the width of neural tuning curves has a large impact on the overall performance of a decision maker (McMillen & Behseta, 2010).

Tuning curves are ubiquitous in neural responses to stimuli (see, e.g., Butts and Goldman (2006)). The relationship between tuning curve shape and decision making performance has intrigued researchers for many years (cf. Pouget, Deneve, Ducom, & Latham, 1999). Naively, one may suppose that task participants improve their performance by sharpening the tuning curves of the neurons involved. However, wider tuning curves are in some cases more efficient in conveying information, and the most informative tuning curve shape depends strongly on the noise distribution in the neurons involved in a particular task (Seriés, Latham, & Pouget, 2004; Zhang & Sejnowski, 1999). Moreover, in several tasks a participant may improve his performance without significantly altering the shapes of the tuning curves in the neurons involved. For instance, in an orientation discrimination task, monkeys are able to learn to discriminate between finer angles over time, while the tuning curves in primary sensory cortex are altered very little (Ghose, Yang, & Maunsell, 2002; Law & Gold, 2008). This suggests that improvements in performance may take place in a learning process downstream from the sensory neurons. (However, see Pilly, Grossberg, and Seitz (2010) for a recent discussion of this issue.)

In this paper we explore the ways in which a participant may improve performance in decision tasks, given tuning curve shapes in sensory neurons. We do not consider the alteration of receptor units' tuning curves, but rather how the information in tuning curves can be utilized more efficiently over the course of many trials. In particular, we are interested in how well participants can do in tasks in which they exhibit a speed-accuracy tradeoff (SAT). Generally, in such tasks, spending more time deciding leads to fewer errors, but at the expense of making fewer decisions per allotted time. With stationary task parameters, the optimal strategy is the one that results in the most correct decisions per unit time: i.e. the one that maximizes the *reward rate*. For a task in which the time to respond is fixed by the experimenter, this amounts to a strategy that chooses the most likely hypothesis. For tasks in which the participant is free to decide at any time, the participant must set his or her own criteria for deciding. We refer to these two types of experimental protocols as the *interrogation* and *free response protocols*, respectively.

Here we are mainly interested in the free response protocol. In the case where the participant is free to respond at any time, and the next trial begins a time D after a response, called the response–stimulus interval (RSI), the reward rate (RR) is defined in terms of the error proportion (ER) and mean reaction time (MRT):

$$\mathrm{RR} = \frac{1 - \mathrm{ER}}{D + \mathrm{MRT}}. \tag{1}$$

The optimal SAT depends on the RSI D (cf. Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006). This may be seen by noting that if D is zero, then the best strategy is to decide as quickly as possible, since even if the ER is what would be achieved by random guessing, the MRT will be small, and the RR will be large. Conversely, if D is very large, one should take one's time to ensure a small probability of error, since there will be few chances for reward in any finite amount of time. Between these two extremes of low accuracy (small MRT) and high accuracy (large MRT) is a range of SATs that are optimal for different values of D. The best procedure, in terms of optimizing RR over the whole range of possible D's, is the one that minimizes the MRT for a given ER. Human participants show evidence of being able to implement this procedure in the case of two-alternative tasks (Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010; Simen et al., 2009). In this paper, we discuss the difficulties inherent in implementing the truly optimal test in the case of more than two alternatives (Brown, Steyvers, & Wagenmakers, 2009), but show how such a test can nevertheless be approximated in a neural network that applies fixed response thresholds to accumulated evidence. The question of where to set those decision thresholds in order to achieve the optimal ER value for a given D is left for future work. In the Discussion, however, we note that it may be possible to generalize an existing procedure for threshold optimization in 2-choice tasks to the case of $N > 2$.

The canonical perceptual decision making task that we will use to illustrate the proposed learning algorithm involves classifying the direction of motion of a visual stimulus into one of several categories. In this task, a participant observes a collection of moving dots on a screen. A certain proportion of them are moving in one of N directions, while the rest are moving randomly. The observer must then determine the direction of motion of the coherently moving dots. In a typical experiment with monkeys as participants, the animal indicates the direction of movement by moving its eyes to a choice target. (See Churchland, Kiani, and Shadlen (2008) and Law and Gold (2008) for recent results in such tasks, and also Niwa and Ditterich (2008) for results in similar tasks using human participants.) The difficulty of this task depends essentially on three factors: (1) the proportion of dots that are moving coherently (the signal-to-noise ratio), (2) the number of possible directions of coordinated movement, and (3) the distance between these alternatives. (With the interrogation protocol, the duration of the stimulus presentation is also a critical factor, with shorter viewing times creating greater difficulty.)

The remainder of this paper is organized as follows. In the following section we describe a network for decision making for the above task. It consists of three layers: an input signal layer, a layer of leaky, competing accumulators (LCA's), and a response output layer. The main result of this paper is a learning algorithm for the weights between the layer of accumulators and the response layer. McMillen and Behseta (2010) showed that overlapping of signals corresponding to different alternatives can be an advantage if the resulting output of the accumulators is multiplied by a matrix that encodes the possible alternatives. This matrix multiplication is achieved by tuning the weights from the accumulator layer to the decision layer so that the vector of weights to each decision unit has the same shape as the vector of signals corresponding to that decision. In Section 3 we describe a simple Hebbian learning rule that performs remarkably well at learning the weights corresponding to near-optimal performance as described in McMillen and Behseta (2010), and we demonstrate analytically why the average weights must converge to them. In Section 4 we present numerical simulations of the blocks of trials involving four alternatives. We see how performance improves over the course of trials, and how the weights corresponding to near-optimal performance are learned for a variety of different shapes for sensory tuning curves. Finally, we conclude in Section 5 with a few remarks and a discussion of how the algorithm we propose can be adapted to situations in which the delay between trials is changing. In this section we also discuss how our work relates to current research trends.

2. The three layer MT-LIP-SC model for decision making

2.1. Overview of the three layers

We propose a three layer neural model for decision making (defined in Table 1, and depicted in Fig. 1). The first layer acts simply as a simple, sensory detector with a preferred motion orientation; the next layer integrates the information from the first layer, but also exhibits competitive dynamics that gradually build

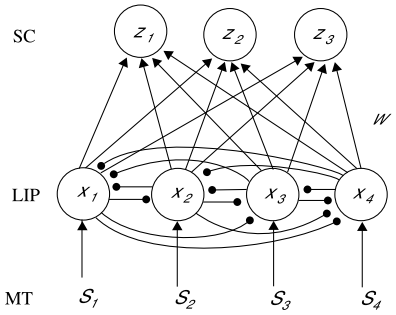

Fig. 1. Neural network model with 4 accumulators and 3 alternatives. The weight matrix W denotes the weights of the connections between the x_i's and z_j's. Arrows represent excitatory connections and circles represent inhibitory connections.

a commitment to one course of action over the alternatives; the last layer triggers a discrete motor response when commitment to one response is sufficiently strong. For convenience, we refer to these three layers, respectively, as the MT, LIP and SC layers. These labels reflect the fact that our model exhibits known properties of neurons in the monkey middle temporal area (MT), the lateral intraparietal (LIP) and the superior colliculus (SC) in decision making tasks requiring eye movements in response to visual motion (Rorie, Gao, McClelland, & Newsome, 2010; Shadlen & Newsome, 2001). (See also Grossberg and Pilly (2008) for a recent discussion of the roles of various parts of the brain in decision making.) The architecture of this circuitry is expected to apply without major modification to other stimulus and response types, however.

We suppose that MT neurons have tuning curves that are preferentially sensitive to a single, given direction of visual motion, and that another layer is stimulated by the activity in this input layer. In this paper we suppose that the tuning curves of the MT neurons do not vary over successive trials. As mentioned above, the tuning curves of sensory neurons may not change much, if at all, over the course of many trials, while performance improves. Here we are interested in how a subject improves performance without any changes to the tuning curves of the sensory neurons. By virtue of their excitatory connections to LIP, model MT units' tuning curves and their excitatory feedforward connections to LIP in turn define tuning curves for LIP units. Questions of major importance in computational neuroscience are: Through what sort of learning process do these tuning curves arise? Can we define an optimal connection scheme that maximizes some function, such as the rate of reward earned by the model? We attempt to make progress on these questions while making the simplifying assumptions that the brain circuits in question are approximately linear systems (at least over a limited range of inputs), and that they employ simple learning schemes (such as Hebbian learning, or error-updating rules such as the Widrow–Hoff, Rescorla–Wagner or delta rules).

Recent work (e.g. McMillen and Holmes (2006) and Bogacz and Gurney (2007)) that avoids discussion of tuning curves shows that under these assumptions a simple neural network can perform approximately optimal hypothesis testing. We now demonstrate that a model consistent with these assumptions does remarkably well at approaching optimal (reward-maximizing) performance in decision making tasks with multiple alternatives. The model's layers are represented mathematically by S, x, and z. Fig. 1 shows a diagram of the model. Notice that, since we take the tuning curves of the MT as given, we do not need to represent the responses of these neurons individually, but only their actions as signals to the LIP layer. Thus, the units in the MT layer represent not necessarily individual MT neurons, but the total weighted sum of signals to a unit in the LIP layer.

Upon presentation of a stimulus, MT neurons present a vector of signals to accumulators in the LIP layer. The signal presented to the

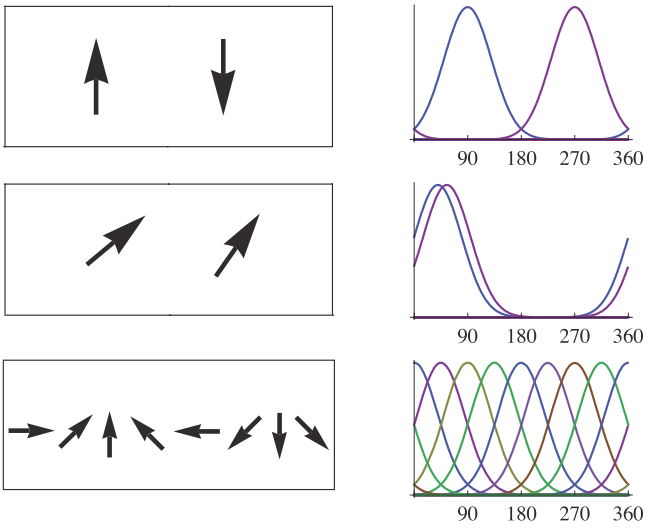

Fig. 2. Possible directions of coherent motion (left panels) and corresponding signal vectors (right panel).

LIP layer is referred to as S_i, representing the total weighted sum of MT signals to the ith accumulator. Each stimulus corresponds to a unique signal, so that the set of signals to the LIP layer may be represented as a vector indexed by μ:

$$S^{\mu} = \left(S_1^{\mu}, S_2^{\mu}, \ldots, S_n^{\mu}\right). \tag{2}$$

The task is to determine which of N possible decision alternatives corresponds to this signal vector. Notice that the size of the signal vector can be greater than the number of decision alternatives, i.e. in general $n > N$.

Although it is not required, we will generally take the S^{μ} signals to be Gaussian:

$$S_i^{\mu} = a \exp\left[-\frac{(i - \mathrm{dir}_{\mu})^2}{2\phi^2}\right], \quad i = 1, \ldots, n. \tag{3}$$

Here dir_{μ} is the peak of the signal, a is the height of the peak, and ϕ is the width of the curve. Notice that if $\phi = 0$, then

$$S_i^{\mu} = a\, \delta_{\mathrm{i,dir}_{\mu}}, \tag{4}$$

where $\delta_{i,j}$ is the Kronecker delta (i.e., 1 if $i = j$, 0 if $i \neq j$), so that the signal is concentrated in the channel dir_{μ}. But, if $\phi > 0$, the signal will have a spread around the peak. Tuning curves associated with the dots motion task have been measured to have a width of about 40° (Law & Gold, 2008). The situation is illustrated in Fig. 2. Directions far apart have very little overlap in the signals, but when the directions are close the overlap is substantial. For the two-alternative case in which dots travel either up or down, the signals have very little overlap. Signals for alternatives corresponding to more similar motion directions have more overlap.

We model the LIP layer as a set of n leaky competing accumulators. The linearized model for their evolution is a stochastic differential equation (Bogacz et al., 2006; McMillen & Holmes, 2006; Usher & McClelland, 2001):

$$dx_i = \left(-kx_i - m\sum_{j \neq i} x_j + S_i\right) dt + c\, dW_i, \quad i = 1, \ldots, n, \tag{5}$$

where k is the decay rate, m is the mutual inhibition, and W_i is a Wiener process (white noise) representing the noise in the signal and from other sources. The noise seen by the accumulators is assumed here to be uncorrelated. That is, the Wiener processes W_i are independent. The signal-to-noise ratio is the ratio of the magnitude of the largest signal to the standard deviation of

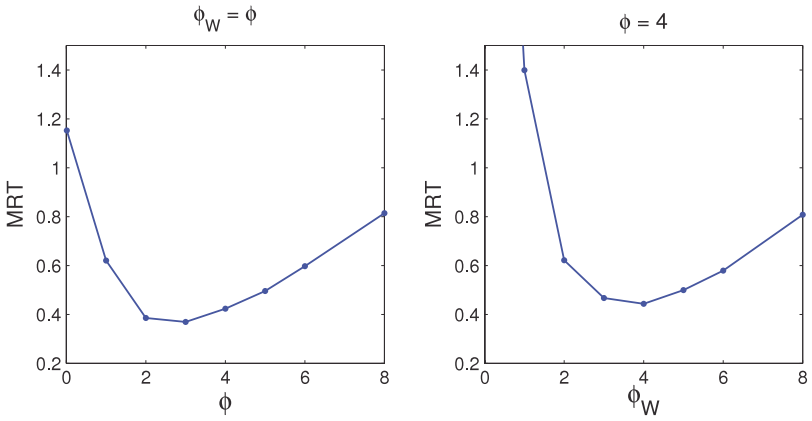

Fig. 3. Effects of signal spread and weight shape. Left panel: MRT vs. spread in the signal vectors, where the weights have the same shape. Right panel: MRT vs. spread in shape of weights with signal vector fixed with $\phi = 4$. In all cases the threshold is such that ER = 0.1.

the noise, i.e. a/c. We can thus model changes in the direction coherence by changing this ratio. The effect of decay and inhibition is to concentrate the values of the accumulators onto the signal vectors. Thus, moderate values of m and k tend to increase the accuracy. Best results are achieved when decay and inhibition are balanced, i.e. $m = k$ (McMillen & Holmes, 2006). For simplicity, and to be concrete, throughout the rest of this paper we will present results for $k = m = 0.5$, $a = 2$ and $c = 1$. Results are qualitatively insensitive to these choices.

The output from accumulator j is fed into the ith unit of SC with weight w_{ij}. Thus, the SC units are given by

$$z_i = f(y_i), \qquad y_i = \sum_{j=1}^{n} w_{ij} x_j. \tag{6}$$

That is, y_i is a weighted sum of the accumulators, and the values z_i of the SC units are obtained by passing these weighted sums through a function $f(x)$. A decision is made once one of the z_i's saturates, or crosses a threshold. We suppose that the functions $f(x)$ are step functions, so that a response is made when SC unit j transitions from 0 to 1 (i.e., when $y_j = \sum_{j=1}^{n} w_{ij} x_j > \theta$). A method for implementing approximate step functions in bistable neural network units with sigmoidal activation functions and strong recurrent feedback is discussed in Simen and Cohen (2009).

The results in this paper are generally applicable, but to be precise, unless stated otherwise, we consider a motion direction task with 36 accumulators and interpret these as representing increments of 10°. If the direction $j \cdot 10°$ is presented, the signal vector takes the shape S_i^{μ} as in (3), with $\mathrm{dir}_{\mu} = j$. For concreteness we consider four possible directions of motion: 30°, 60°, 140°, 220°. Thus, if, say, the direction of coordinated movement is 60°, the signal vector has a peak at the sixth accumulator. The four possibilities are represented by the four possible signal vectors with peaks at accumulators 3, 6, 14 and 22. In this paper we only consider the case when all the possibilities are equally likely, in which case the appropriate initial condition for the accumulators is $x_i(0) = 0$.

2.2. Optimality and weights

McMillen and Behseta (2010) showed that approximately optimal performance is achieved when the weights mimic the shape of the possible incoming signal vectors. That is to say, a threshold crossing test best approximates the optimal test when $w_{ij} \propto S_j^{\mu_i}$. The magnitude of the weights are not important in terms of optimality, as the magnitude may be incorporated into the thresholds.

For two alternatives the sequential probability ratio test (SPRT) minimizes reaction times among all tests making a decision with a given error rate, and hence optimizes reward rate (Wald & Wolfowitz, 1948). For more than 2 alternatives, there is no single optimal test in the sense the SPRT is optimal. However, the multihypothesis sequential probability ratio test (MSPRT) is an asymptotically optimal test that achieves the smallest MRT among all tests for a given ER, in the limit as the ER approaches zero (Dragalin et al., 1999). In the context of the three layer model presented here, the asymptotically optimal test is the one in which the weights $w_{ij} \propto S_j^{\mu_i}$, and a decision in favor of hypothesis i is made when the quantity

$$y_i - \max_{i \neq j} y_j \tag{7}$$

crosses a threshold. For this reason, the asymptotically optimal test is referred to as the max-vs-next test. A related, but sub-optimal, test is the max-vs-average test, which chooses hypothesis i when the quantity

$$y_i - \frac{1}{N-1} \sum_{j \neq i} y_j \tag{8}$$

crosses a threshold. The max-vs-average test may be described by the somewhat cumbersome term,"nearly asymptotically optimal". That is, it is nearly as good as the asymptotically optimal test. The advantage of the mutual inhibition term in (5) is that a simple threshold crossing test (i.e. choose i when z_i crosses a threshold) achieves results indistinguishable from a max-vs-average test when the inhibition term is of moderate size. For a fuller discussion of the relationship between these sub-optimal tests, we refer the reader to McMillen and Holmes (2006).

In this paper we consider only the simple threshold crossing tests, which approximate the sub-optimal max-vs-average test. This restriction is made since, (a) it is simple to implement (it requires only a consideration of the weighted accumulator values), and (b) here we are mainly interested in how a network may learn to approximate optimality. For this reason, for the remainder of this paper, we will use the term "optimal" in the sense that the three layer network with a simple threshold crossing test can achieve the best possible reward rate. Thus, we consider a constrained optimality problem.

The performance of the threshold crossing tests is illustrated in Fig. 3. Here we consider a test with 36 accumulators and the four alternatives as described above. In Fig. 3 we plot the MRT for a fixed value of the ER (ER = 0.1). For each value of the spread we compute the threshold such that ER = 0.1, and find the corresponding MRT. Each panel demonstrates an important fact, as we elucidate below.

In the left panel of Fig. 3 we take the signal vectors to be as in (3), and allow ϕ to vary. Thus, $\phi = 0$ corresponds to the case

when the signal is concentrated in a single channel. Positive values of ϕ correspond to signals that are spread about a peak. In these computations, the weights are the optimal weights, i.e. $w_{ij} \propto S_j^{\mu_i}$. This panel thus shows the minimal MRT that can be achieved by a threshold crossing test for an ER of 0.1. We see that there is an advantage to a moderate spread in the signals if this information can be utilized by the decision mechanism. In fact, the minimal MRT occurs when the spread is near $\phi = 3$. It is interesting to note that this corresponds to a width in the shape of the signal vectors of about 30°, while the width of tuning curves in MT associated with the direction task as measured in Law and Gold (2008) is approximately 40°.

In the right panel we fix the spread in the signal vectors at $\phi = 4$, and compute MRT for various spreads in the weights. In order to get an idea of how the spread in the shape of the weights affects performance when the signal shape is fixed, in these simulations we suppose that the weights also have a Gaussian shape:

$$w_{ij} = w_0 \exp\left[-\frac{(i-j)^2}{2\phi_W^2}\right], \quad j = 1, \ldots, n, \tag{9}$$

where w_0 is a normalizing factor chosen so that $\sum_{j=1}^n w_{ij}^2 = 1$. The spread ϕ_W controls how the values of the accumulators are weighted before making a decision. A value $\phi_W = \phi$ corresponds to the optimal weights. In the case $\phi_W = 0$, we have $y_i = x_i$, so that the accumulator values are not weighted. When $\phi_W = \infty$, each y_i is the same, i.e. the sum of all accumulators. The right panel of Fig. 3 shows that MRT is minimized when $\phi_W = \phi$. That is, the best results are achieved when the weights are the optimal weights. In this figure both the signal vector and the weight vectors have Gaussian shapes. MRT is minimized (and hence the reward rate is maximized) when the widths of both are the same. We have performed a number of simulations comparing results for various shapes of the signal and weight vector, and have found in all cases that the MRT is minimized when the shape of the weight vector is the same as the signal vector. See McMillen and Behseta (2010) for a detailed discussion. For this reason, we will refer to weights that have the same shape as the possible signal vectors as *optimal weights*, keeping in mind that these are the weights that solve the constrained optimality problem we have described.

To reiterate, a moderate spread in the signals is a significant advantage, but only if the LIP-to-SC weights can be tuned to the MSPRT-implementing weights, that is, to take on the same shape as the possible signal vectors defined by MT activity. In the following section we consider how the weights may be modified over the course of trials.

3. An algorithm for learning the LIP-to-SC weights

We propose a simple Hebbian weight learning algorithm for the weights w_{ij}. The learning algorithm is a modification of the classical Widrow–Hoff rule. The theory of this update rule is described, e.g. in Hertz, Krogh, and Palmer (1991). In rules of this type, the connection strength being modified acts as a filter that tracks an input signal. At any point, its value is an exponentially decaying, time average of past input values. High frequency changes in this signal (representing noise) are filtered out by the algorithm, producing little change in the updated weight. In contrast, low frequency signal changes (representing, hopefully, the uncorrupted input signal) produce significant changes in the weight. If the signal is constant and noise is absent, the weight will converge exponentially on the value of the signal. If what is being tracked is a signal that depends on the product of activations in a sending unit and a receiving unit, then this rule is simply a Hebbian update rule with a decay term for forgetting old co-activation levels — a useful feature in a noisy neural system.

After each trial the participant responds with a choice among alternatives, say i. At this time the weights to the output unit z_i corresponding to the choice made are updated, according to whether a reward is received or not. Then, if the choice corresponding to z_i is chosen, the weights are updated by the rule

$$w_{ij}^{\text{new}} = (1-\alpha) w_{ij}^{\text{old}} + \alpha\, \Delta w_{ij}, \tag{10}$$

$$\Delta w_{ij} = r z_i x_j, \tag{11}$$

where r is the magnitude of the reward, and α is the learning rate. Notice that only the weights to the unit corresponding to the choice made are updated, and this is the sense in which the rule is Hebbian. In this discrete algorithm, we suppose that a reward is either earned or not so that r is either 1 or 0 depending on whether a correct decision is made. This rule can easily be modified to take into account a probability of receiving a reward for a correct response, but we do not consider such modifications here. After each trial, the weights are normalized so that

$$\sum_{j=1}^n w_{ij}^2 = w_0. \tag{12}$$

We usually take $w_0 = 1$. The normalization (12) is thought to be a common feature of synaptic plasticity (Royer & Pare, 2003). Note that this normalization means that if an incorrect decision is made, then the weights are unchanged, since then $\Delta w_{ij} = 0$, and the normalization will cancel the multiplication by $1 - \alpha$.

Thus, after each trial, if a correct decision is made the weights to the correct output unit are increased in proportion to the values of the accumulators $\mathbf{x}$. There is no need to estimate the probability of making a correct decision or an expected value of the reward, as in reinforcement learning methods based on prediction errors (e.g., Law & Gold, 2009), since only the values of the units are used in the update rule. With this rule the weights track the shape of the vectors being passed from the LIP layer. The weights thus tend to oscillate around the means of the accumulator values, $\langle x_j(t) \rangle$.

We now prove that the accumulator values on average take on the shape of the signal vector from the MT layer. The first step is to write Eq. (5) as

$$dx_i = \left(\lambda x_i - m \sum_{j=1}^n x_j + S_i\right) dt + c\, dW_i, \quad i = 1, \ldots, n, \tag{13}$$

where $\lambda = m - k$ is the difference between inhibition and decay. Since the mean of the Ito integral $\int_0^t dW_i(\tau) d\tau$ vanishes (Gardiner, 2004), the mean value of the accumulators obey the ordinary

Table 1
Three layer model with weight learning rule. Numbers at left refer to equations in the text.

	$S_i,\ i = 1, \ldots, n$	Given signals from MT
(5)	$dx_i = \left(-k x_i - m \sum_{j \neq i} x_j + S_i\right) dt + c\, dW_i$	Accumulator evolution in LIP
(6)	$z_i = f(y_i),\ y_i = \sum_{j=1}^n w_{ij} x_j$	Decision units in SC
(10)	$w_{ij}^{\text{new}} = (1-\alpha) w_{ij}^{\text{old}} + \alpha\, \Delta w_{ij}$	LIP-to-SC weight learning rule
(11)	$\Delta w_{ij} = r z_i x_j$	
(12)	$\sum_{j=1}^n w_{ij}^2 = w_0$	Weight normalization

differential equation

$$\frac{d}{dt}\langle x_i(t)\rangle = \lambda\langle x_i(t)\rangle - m\left\langle\sum_{j=1}^{n} x_j(t)\right\rangle + S_i. \tag{14}$$

We calculate the mean of the sum of the accumulators as follows by summing the Eq. (5) to deduce that the increment of the sum of the accumulators is

$$d\left(\sum_{j=1}^{n} x_j\right) = \left(-(k+(n-1)m)\sum_{j=1}^{n} x_j + \sum_{j=1}^{n} S_j\right)dt + c\sum_{j=1}^{n} dW_j. \tag{15}$$

Since the sum of Wiener processes has mean zero, the mean of the sum obeys the ordinary differential equation

$$\frac{d}{dt}\left\langle\sum_{j=1}^{n} x_i\right\rangle = -\hat{\lambda}\left\langle\sum_{j=1}^{n} x_j\right\rangle + \sum_{j=1}^{n} S_j, \tag{16}$$

where $\hat{\lambda} = k + (n-1)m$. Therefore,

$$\left\langle\sum_{j=1}^{n} x_j(t)\right\rangle = \sum_{j=1}^{n} S_j\frac{1-e^{-\hat{\lambda}t}}{\hat{\lambda}}. \tag{17}$$

Substituting the above expression into (14), we obtain an ordinary differential equation for $\langle x_i(t)\rangle$. Solving this, we obtain

$$\langle x_i(t)\rangle = \left(S_i - \frac{1}{n}\sum_{j=1}^{n} S_j\right)\frac{e^{\lambda t}-1}{\lambda} + \frac{1}{n}\sum_{j=1}^{n} S_j\frac{1-e^{-\hat{\lambda}t}}{\hat{\lambda}}. \tag{18}$$

Notice, now, that the mean of $x_i(t)$ has the form

$$\langle x_i(t)\rangle = \left(S_i - \frac{1}{n}\sum_{j=1}^{n} S_j\right)c_1(t) + c_2(t), \tag{19}$$

where c_1 and c_2 are functions of t, but they do not depend on i. Notice that c_1 and c_2 are strictly increasing functions of t. Thus the mean of the vector of accumulators tends toward the shape of the vector of signals from MT, amplified by the function $c_1(t)$. Thus, the accumulators tend to sharpen around a peak corresponding to the largest signal. The behavior of the means of the accumulators and the role of inhibition may be better understood by considering the case in which decay is nearly balanced with inhibition, i.e. $k \approx m$. Then $\lambda \approx 0$ and $\hat{\lambda} \approx nm$, and from (18), the means of the accumulators are

$$\langle x_i(t)\rangle \approx \begin{cases} \left(S_i - \frac{1}{n}\sum_{j=1}^{n} S_j\right)t & \text{if } m \gg 0 \\ S_i\, t & \text{if } m \approx 0. \end{cases} \tag{20}$$

The effect of inhibition may be clearly seen from the above. With no inhibition, the accumulators evolve independently, simply amplified over time. When inhibition is present, those accumulators corresponding to signals that are less than the average trend negatively, and the difference between the largest and smallest accumulator is thus amplified. It is worth noting that lateral inhibition causes the accumulators to be non-independent stochastic processes. This does not affect the network's approximation of the MSPRT, however; see McMillen and Behseta (2010) and McMillen and Holmes (2006).

The update rule (10)–(11) causes the weights to track the normalized mean values of the accumulators and thus causes the weights to track values whose means take on the shape of the MT-to-LIP signal vectors. In the extreme cases of (20) where $\hat{\lambda}$ is either large or small, the weights will thus tend to either $S_i - \frac{1}{n}\sum_{j=1}^{n} S_j$ or S_i, both of which obviously have the same shape as the signal vector. Notice that over time, due to the amplification by $c_1(t)$, the vector of accumulators will generally become scaled up versions of the vector of signals. This scaling up is a sharpening in the sense that the ratio of the height of the integrated signal to the variance of the noise becomes larger over time. However, this is not a sharpening in the sense in which a probability density sharpens by increasing its peak while integrating to one. In general, as we demonstrate in simulations below, the weights track these scaled up versions of the signal vector. The degree of the scaling depends on the time of integration, the MRT on average. That is, the weights tend, on average, to mimic the shape of the signal vector with oscillations about this shape that depend on the learning rate.

Fig. 4 shows the dynamics of evidence accumulation within trials, demonstrating that Gaussian bumps of activation arise on the LIP layer (preserving the Gaussian input signal profiles, and therefore producing Gaussian LIP-to-SC weights through Hebbian learning; see panel A). The bottom of panel A shows how the weighted LIP activations (black asterisks) approach or recede from the thresholds (red asterisks) implemented by each of three SC units. Panel B demonstrates that the weighted sum of LIP activations in favor of choice 2 (blue solid) increases more rapidly on average than the most sensitive LIP unit alone (blue dashed). This occurs without any appreciable increase in noise, and therefore results in better performance than basing decisions on evidence from a single LIP unit alone. In panel C, the same situation occurs, but a threshold is now applied to the evidence in order to produce a decision. Red and blue traces fall off over time because the average is based on fewer and fewer trials as time progresses (more and more decisions have already taken place by the end of the plot).

4. Results of simulations

Fig. 5 shows results of simulations using the update rule (10)–(11) in the free response protocol. The weights are initially chosen randomly, with a peak added at w_{ii}. We see how the weights evolve over time, and how this affects the performance of the participant. The reward rate continually increases on average, and the ER continually decreases. The bottom panels show the weights to SC corresponding to $i = 14$, or to angle 140°. The weights for the other alternatives behave similarly. Simulations in which the weights are chosen differently show similar improvements in performance and similar matching of the weight profiles to the signal vector shapes. Cases in which the weights are all chosen randomly show a more dramatic improvement in RR since then the accuracy will initially be very low. Fig. 5 shows that even when the weight has a peak at the right position, a dramatic improvement occurs: for example, the RR more than doubles and the RT and ER both decrease over time.

Fig. 5 shows one block of 500 trials. In order to see how the weight update rule behaves on average, we carried out the same simulation for a number of blocks and then averaged the weights over trials in each block, and then took the average over 150 blocks of trials. The values of such averaged weights are seen in Fig. 6. In this figure we show the averaged weights for different values of the threshold, as well as different values of the spread in the signals. We see that on average, the weight profile is close to the signal shape. Also shown in these figures are the ERs and MRTs for these blocks of trials. Notice that in the lower left panel, the ER = 0.58 is not much smaller than would be achieved by random guessing (0.75). In this case the threshold is very small, as is the corresponding MRT of 0.09. In this situation it will take the weights much longer to learn the shape of the signal vectors, since most of the time the decision will be incorrect. This is why the weights appear more erratic in this frame than in the others. However, even

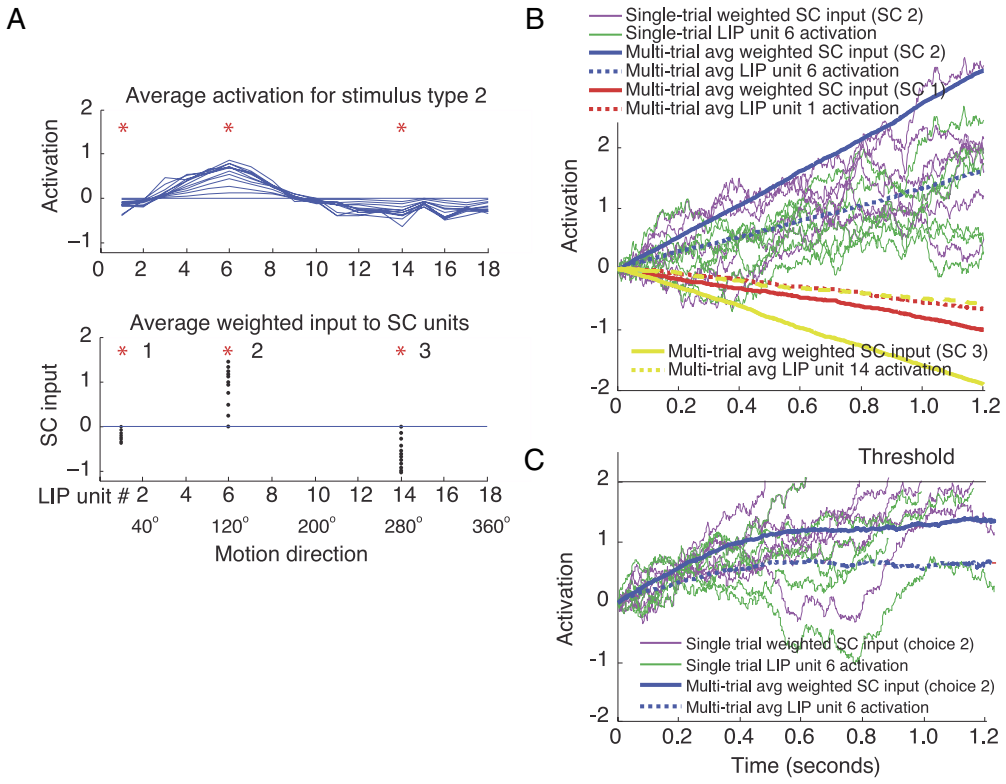

Fig. 4. Panel A, top, shows LIP unit activations at several time points within a decision. Activations are averaged over many instances of stimulus type 2, which produces maximal activation in LIP unit 6 (motion direction 120°; here we arbitrarily quantized directions into 18 levels). Panel A, bottom, shows the weighted values of these activations feeding into each of 3 SC units. Panel B shows the average state of weighted evidence accumulation over multiple trials for input to SC unit 2 (blue solid) and average LIP unit 6 activity (blue dashed) without applying thresholds applied to the evidence (the interrogation protocol). Here, red and yellow indicate the weighted evidence in favor of the other two responses, corresponding to stimuli with maximal activation at units 1 and 14 (10° and 200°, solid and dashed respectively). Red and yellow dashed lines indicate activation in the LIP units most sensitive to these stimuli. Individual timecourses of activity for 10 trials are shown for the weighted input to SC 2 (magenta) and for LIP 6 (green). Panel C shows the average state of weighted evidence accumulation for SC 2 (blue solid) and average unit 6 activity (blue dashed) within free response trials (black line indicates threshold). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

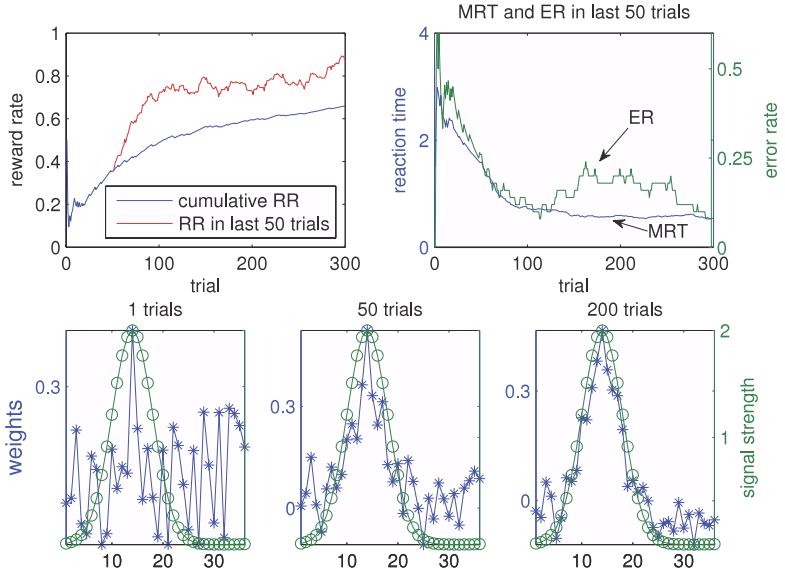

Fig. 5. Effects of weight learning rule. The threshold is fixed at $\theta = 2$. There are four alternatives (3, 6, 14, 22), and the learning rate is $\alpha = 0.01$. In the bottom panel the signal strength is plotted on the right axis (circles), and the weights are shown on the left axis (stars). The RSI used in the calculation of RR is $D = 0.5$.

in this case, the average values of the weights take the same shape as the signal vector. Similar comments apply, *mutatis mutandis*, to the upper left panel.

We also performed simulations using different shapes for the signal vectors, in order to verify that the Gaussian shape achieved by the weight vectors is not simply an artifact of the Gaussian noise in the equation for the LIP units. This is seen in Fig. 7. Here we plot the averaged weights using cosine and $1 - |x|$ shapes for the signal vectors. As with the Gaussian shapes, we see that the weight vectors tend, on average, to the shape of the signal vectors.

Additionally, we performed a variety of simulations to examine the effects of the various parameters in the model. Generally, the

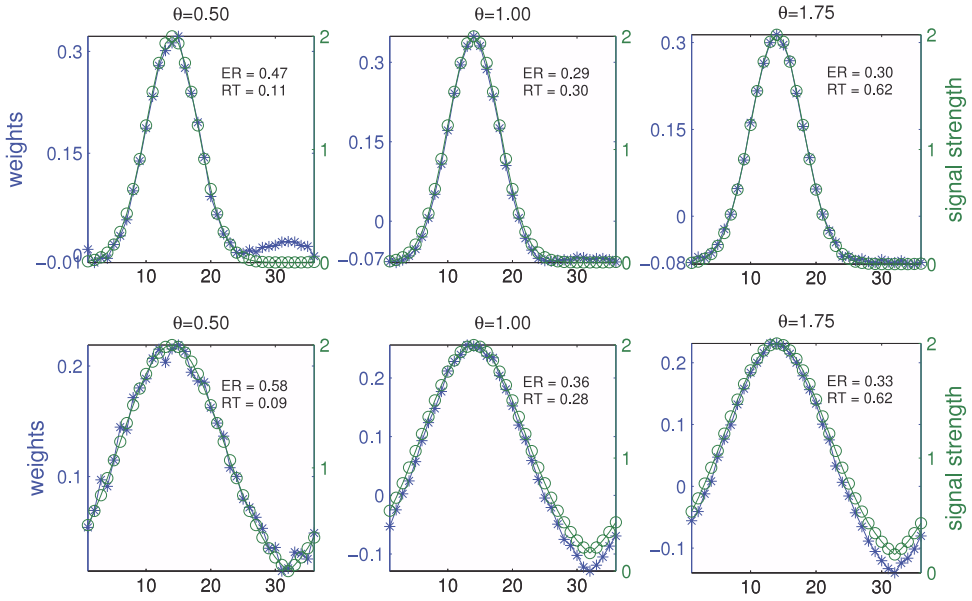

Fig. 6. Averaged weights over 150 blocks of 500 trials. In the top row $\phi = 4$; in the bottom row $\phi = 8$.

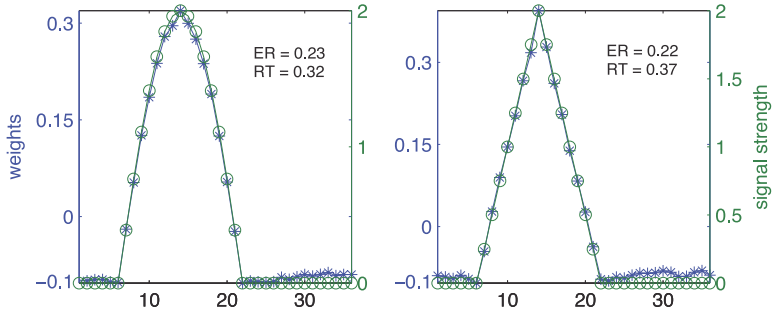

Fig. 7. Averaged weights with cosine and $1 - |x|$ shapes.

model is insensitive to changes in the parameters a, c, k, w, in the sense that the weights tend on average toward the optimal weight shape mimicking the shape of the signal vectors. If the learning rate α is made smaller, the weights take longer to track to the shape of the signals, but there is less variation around these mean values.

5. Discussion

5.1. Comments on the present work and directions of future research

The simple rule (10)–(11) works remarkably well at learning the shapes of the signal vectors from MT to LIP. This leads to a dramatic improvement in performance, and occurs without any direct connection to the MT layer. The three layer model incorporates integration of information, a rule for making the decision, as well as a simple algorithm for learning to optimize reward rates by learning the shapes of the vectors of neural signals coming from an input layer. These components are the essential aspects of a complete decision-theoretic model.

A great advantage of the simple model we have proposed is its flexibility. For instance, one can take a continuous limit in the update rule (10)–(11). The Eq. (10) may be rewritten as

$$\frac{w_{ij}^{\text{new}} - w_{ij}^{\text{old}}}{\alpha} = -w_{ij}^{\text{old}} + r z_i x_j. \tag{21}$$

Therefore, supposing that the weights w_{ij} are functions of time and that w_{ij}^{new} is the update to the weight w_{ij}^{old} after a time step of α, we may take the limit as this time step approaches zero to obtain the continuous version of the update rule:

$$\frac{dw_{ij}(t)}{dt} = -w_{ij}(t) + r(t) z_i(t) x_j(t). \tag{22}$$

Notice that if the threshold θ for the y_i's is fixed, an overall increase in the weights w_{ij} has the effect of amplifying the accumulators. Thus, the threshold is crossed sooner, and the effective threshold is thereby reduced. The continuous version of the update rule (22) can thus be modified to a scheme that not only updates the shapes of the weights, but the normalizing factor w_0 of the weights, and hence the overall strength of the weights.

Such an algorithm was used in the case of two alternatives in Simen and Cohen (2009). (See a related algorithm in Simen, Cohen, and Holmes (2006).) In the two-alternative context in which different responses are rewarded at random intervals, but with different expected delays to reward, the model selects the more frequently rewarded response more often (the ratio of responses of each type in fact matches the ratio of rewards earned for each response type, consistent with the 'matching law' of behavioral psychology (Herrnstein, 1997)). In a work in progress we examine a modified version of this approach that not only updates the thresholds but the weight shapes defining a tuning curve as well. Thus the continuous-time version of the algorithm already has a promising connection to well-known behavioral findings.

The model is amenable to modifications to account for changing conditions across trials. Reward rates are optimized over all

networks with the structure depicted in Fig. 1 when the weight shapes are fixed at the shapes of the possible signal vectors. However, the rule (10)–(11) incorporates a variation around the means of these signals if the learning rate $\alpha > 0$. In a sequence of trials, one may reduce α over time so that these variations continually decrease. In effect, one learns the incoming signals by first allowing α to be moderately large, and then reducing α as the weights get closer and closer to the means of the accumulators in the LIP layer.

A number of issues raised by this work remain to be explored. One of these is to see how this model may be modified to account for situations in which various aspects such as the RSI, number of alternatives and signal-to-noise ratio are changing from trial to trial. We have also assumed that the signals from MT to LIP are constant across trials, which is obviously not realistic. Hebbian learning rules such as the one considered in the present paper may be employed to learn weights from MT to LIP neurons, and thus adjust to different tasks. A temporary increase in α, for example, could be used to adapt to any sudden changes in task conditions.

5.2. The place of the present work in the context of recent research

The network we propose has several advantages. For one, it is extremely simple, as can be seen in the summary of the complete model in Table 1 in Section 3. This simplicity makes the model tractable for analysis, and it also suggests an extremely simple physical substrate that could plausibly be implemented in the brain (a substrate composed of the same electric circuit components used to model the individual neuronal membrane). The model also incorporates decision making criteria: a decision is made once one of the units in the decision layer crosses a threshold. Thus we have focussed our discussion on how it performs in the free response protocol. It is trivial to adapt the model to the interrogation protocol: if the weights are correctly tuned, then at any given time the most likely hypothesis corresponds to the one with the largest weighted sum of LIP unit activities (McMillen & Behseta, 2010, cf. Eq. (44)). The network is also easily adaptable. If the signal vectors change, the network can quickly learn the optimal weights to encode these new alternatives.

Furman and Wang (2008) investigated one generalization of such networks in the context of a more detailed but less analytically tractable model, and suggested that a continuous decision was difficult for reduced models of the type investigated here. However, we have shown that the solution to the continuum problem is straightforward: simply adding more units as done by Furman and Wang (2008) is sufficient for achieving decision making in the continuum context. Moreover, when decision making requires a discrete set of responses, nearly optimal performance can be analytically assessed for the LCA model through a direct mapping onto the MSPRT, the only known, computationally tractable N-alternative hypothesis testing procedure with fixed thresholds that approximates optimal performance. The model of Bogacz and Gurney (2007) also leverages the same analytical tractability to implement the MSPRT in the case of Kronecker delta tuning curves.

Beck et al. (2008) investigated the same problem from the perspective of optimal integration of information using a scheme that requires spike statistics and the way that neurons integrate spikes to meet certain restrictions. For the case of multiple alternative decisions, the authors evaluate their theory by comparison with single-cell recordings from monkey lateral intraparietal cortex during the performance of a random-dot motion discrimination task. The optimal evidence accumulation then is obtained via linear integration of neural activity and eventually is modeled with a straightforward Bayesian formulation across trials and over time. The proposed Bayesian model is then used to validate the tradeoff between accuracy and speed in decision making.

Optimality, in Beck et al. (2008), means that the accumulation of evidence is done without loss of information, so that the most likely action is chosen through dynamic attractors. In the context of the interrogation protocol, where the decision is made after a fixed time, this definition of optimality coincides with the one employed in the present paper, of maximizing reward. With their primary focus on information integration, Beck et al. (2008) do not address the issue of when to make the decision, and therefore their approach does not address how reward rates can be maximized in the free response protocol where the participant may decide at any time. The advantage of the model proposed here is that with nothing more than a classic, nearly linear firing-rate model that can be implemented with an economy of physical components, we can implement an approximately optimal decision making procedure and thereby give a complete, decision-theoretic account of decision making by these networks.

Furthermore, the dynamics of decision making in our model are qualitatively the same as in the aforementioned models. On each trial, an initial Gaussian signal that is buried in a white noise background is scaled up vertically, while noise is reduced relative to the peak signal amplitude. The resulting increase in the ratio of the integrated signal to the noise produces greater accuracy the longer the stimulus is viewed. This cleaned up signal is then effectively compared to a template of possible signal alternatives that is learned over many trials before being compared to a threshold for response initiation, and threshold adaptation can then be used to adjust speed-accuracy tradeoffs by the model. Such a process requires no violation of the assumptions made in linear systems theory, and is therefore highly analytically tractable.

The similarity of the activation dynamics arising from these different modeling approaches suggests that the overall shape of a developing neural theory of decision making may be robust to variations in detailed implementation — a point we should find encouraging, since all tractable approaches to brain modeling depend on simplifying assumptions that are bound to be wrong at a more fine-grained level of description.

Given its compatibility with linear systems theory, it is, finally, worth noting that our model is quite compatible with linear models of perceptual learning that learn to pool the information from MT cells to produce efficient evidence integration in LIP (e.g. Law & Gold, 2009). Optimal linear classification of perceptual signals presented for fixed durations was investigated by these authors, and the reinforcement learning process they apply leads to a set of MT-to-LIP weights that would complement the set of LIP-to-SC weights that our learning rule acquires. In future work, we plan to incorporate learning at all levels, and to investigate whether the reward-modulated Hebbian learning approach we take here is as capable of learning linear-optimal decoding schemes as the prediction-error-based approach of Law and Gold (2009) (which is not applicable to free response tasks), and the incremental reweighting approach of Petrov, Dosher, and Lu (2005). In general, our work continues the progression begun by these and other models, which suggest that the neural retuning underlying perceptual learning occurs farther away from the primary sensory cortices than originally believed. The model of perceptual learning we propose suggests that some amount of learning in fact happens at a stage that is still farther from sensory inputs and closer to motor outputs: it can take the outputs of linear-optimal stimulus classification procedures and convert them into linear-optimal evidence accumulation processes that achieve reward-maximizing speed-accuracy tradeoffs.

Acknowledgements

We would like to thank two anonymous reviewers for their helpful comments on an early draft of the manuscript.

References

Beck, J., Ma, W., Kiani, R., Hanks, T., Churchland, A., Roitman, J., et al. (2008). Probabalistic population codes for bayesian decision making. *Neuron*, *60*, 1142–1152.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. (2006). The physics of optimal decision making: a formal analysis of performance in two-alternative forced choice tasks. *Psychological Review*, *113*(4), 700–765.

Bogacz, R., & Gurney, K. (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Computation*, *19*(2), 442–477.

Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences*, *33*(1), 10–16.

Brown, S., Steyvers, M., & Wagenmakers, E.-J. (2009). Observing evidence accumulation during multi-alternative decisions. *Journal of Mathematical Psychology*, *53*(6), 453–462.

Butts, D. A., & Goldman, M. S. (2006). Tuning curves, neuronal variability, and sensory coding. *PLoS Biology*, *4*(4), e92.

Churchland, A., Kiani, R., & Shadlen, M. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, *11*(6), 693–702.

Dragalin, V., Tartakovsky, A., & Veeravalli, V. (1999). Multihypothesis sequential probability ratio tests, part I: asymptotic optimality. *IEEE Transactions on Information*, *45*, 2448–2461.

Furman, M., & Wang, X. (2008). Similarity effect and optimal control of multiple-choice decision making. *Neuron*, *60*, 1153–1168.

Gardiner, C. (2004). *Handbook of stochastic methods* (3rd ed.). Berlin: Springer-Verlag.

Ghose, G., Yang, T., & Maunsell, J. (2002). Physiological correlates of perceptual learning in monkey v1 and v2. *Journal of Neurophysiology*, *87*, 1867–1888.

Gold, J., & Shadlen, M. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences*, *5*, 10–16.

Grossberg, S., & Pilly, P. K. (2008). Temporal dynamics of decision-making during motion perception in the visual cortex. *Vision Research*, *48*(12), 1345–1373.

Herrnstein, R. (1997). *The matching law: papers in psychology and economics*. Cambridge, MA: Harvard University Press.

Hertz, J., Krogh, A., & Palmer, R. (1991). Introduction to the theory of neural computation. In *Lecture notes*: *Vol. 1. Santa Fe institute studies in the sciences of complexity*. Redwood City, CA: Addison-Wesley Publishing Company Advanced Book Program.

Knill, D., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719.

Law, C.-T., & Gold, J. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory cortical area. *Nature Neurosci.*, *11*, 505–513.

Law, C.-T., & Gold, J. (2009). Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nature Neurosci.*, *12*(5), 655–663.

McMillen, T., & Behseta, S. (2010). On the effects of signal acuity in a multi-alternative model of decision making. *Neural Computation*, *22*(2), 539–580.

McMillen, T., & Holmes, P. (2006). The dynamics of choice among multiple alternatives. *Journal of Mathematical Psychology*, *50*(1), 30–57.

Niwa, M., & Ditterich, J. (2008). Perceptual decisions between directions of visual motion. *J. Neurosci.*, *28*(17), 4435–4445.

Petrov, E. A., Dosher, B. A., & Lu, Z.-L. (2005). The dynamics of perceptual learning: an incremental reweighting model. *Psychological Review*, *112*, 715–743.

Pilly, P. K., Grossberg, S., & Seitz, A. R. (2010). Low-level sensory plasticity during task-irrelevant perceptual learning: evidence from conventional and double training procedures. *Vision Research*, *50*(4), 424–432. Perceptual Learning Part II.

Pouget, A., Deneve, S., Ducom, J.-C., & Latham, P. (1999). Narrow vs. wide tuning curves: what's best for a population code? *Neural Computation*, *11*, 85–90.

Rorie, A., Gao, J., McClelland, J., & Newsome, W. (2010). Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. *PLoS ONE*, *5*(2), e9308.

Royer, S., & Pare, D. (2003). Conservation of total synaptic weight through balanced synaptic depression and potentiation. *Nature*, *422*, 518–522.

Seriés, P., Latham, P., & Pouget, A. (2004). Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. *Nature Neurosci.*, *7*(10), 1129–1135.

Shadlen, M., & Newsome, W. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*, 1916–1936.

Simen, P., & Cohen, J. (2009). Explicit melioration by a neural diffusion model. *Brain Research*, *1299*, 95–117.

Simen, P., Cohen, J., & Holmes, P. (2006). Rapid decision threshold modulation by reward rate in a neural network. *Neural Networks*, *19*, 1013–1026.

Simen, P., Contreras, D., Buck, C., Hu, P., Holmes, P., & Cohen, J. D. (2009). Reward rate optimization in two-alternative decision making: empirical tests of theoretical predictions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1865–1897.

Usher, M., & McClelland, J. (2001). On the time course of perceptual choice: the leaky competing accumulator model. *Psychological Review*, *108*, 550–592.

Wald, A., & Wolfowitz, J. (1948). Optimal character of the sequential probability ratio test. *Annals of Mathematical Statistics*, *19*, 326–339.

Zhang, K., & Sejnowski, T. (1999). Neural tuning: to sharpen or broaden? *Neural Computation*, *11*, 75–84.