LETTER **Communicated by Richard Zemel**

Bayesian Spiking Neurons I: Inference

Sophie Deneve
sophie.deneve@ens.fr
Group for Neural Theory, Département d'Etudes Cognitives,
Ecole Normale Supérieure, Collège de France, 75005 Paris, France

We show that the dynamics of spiking neurons can be interpreted as a form of Bayesian inference in time. Neurons that optimally integrate evidence about events in the external world exhibit properties similar to leaky integrate-and-fire neurons with spike-dependent adaptation and maximally respond to fluctuations of their input. Spikes signal the occurrence of new information—what cannot be predicted from the past activity. As a result, firing statistics are close to Poisson, albeit providing a deterministic representation of probabilities.

1 Introduction

Many perceptual and motor tasks performed by the central nervous system are probabilistic in nature, and can be described in a Bayesian framework (Kording & Wolpert, 2004; Knill & Richards, 1996). According to this theory, one probability is assigned to all possible interpretations of the available sensory and motor information on the basis of sensory or motor noise and priors designating the most likely interpretation. The percept, or the motor output, corresponds to the interpretation that is most probable or has the most desirable outcome.

Thus, perception can be considered as an inference process extracting the state of important sensory variables, such as direction of motion, orientation of surfaces, object boundaries, and identities, from a noisy and ambiguous sensory input. The most famous sensory ambiguities include the ill-posed 2D-to-3D transform necessary to infer the 3D structure of objects from their 2D projection on the retina and the aperture problem, where the movement of an edge whose boundaries cannot be seen is compatible with many different directions of motion. The use of Bayesian inference and priors in human perception recently has been shown to account for a wide range of perceptual phenomena (Knill & Richards, 1996; Weiss & Freeman, 2001; Weiss & Fleet, 2002; Feldman, 2001; Geisler, Perry, Super, & Gallogly, 2001; van Beers, Sittig, & Gon, 1999; Ernst & Banks, 2002). Similar to perception, sensorimotor integration and motor control require probabilistic computation: motor effectors are noisy, and motor goals are ambiguous, since many different series of muscle contractions could

reach the same goal (Ghahramani, Wolpert, & Jordan, 1995). Not surprisingly, Bayesian inference and priors also seem to be used by humans in these computations (Kording & Wolpert, 2004; Wolpert & Ghahramani, 2000).

Such probabilistic computations have to be performed as quickly and accurately as possible in a perpetually changing world. This is particularly striking in the motor domain, where the position of the motor effector and the resulting sensory feedback are constantly modified, but it is also true in the perceptual domain, where this temporal dimension is often neglected. Objects move, appear, and disappear, unpredictably. The retinotopic visual input and its interpretation change after each saccadic eye movement, which occurs on average every 250 ms in humans (Ballard, Hayhoe, Salgian, & Shinoda, 2000). Thus, the problem of perception and action is essentially a form of Bayesian filtering: the state of sensory and motor variables has to be estimated online from priors and a stream of noisy and ambiguous observations.

It is essential to understand the neural basis of these computations—how neurons or networks of neurons represent and compute with probabilities and learn probabilistic models. In particular, as these computations have to be performed on a temporal scale of the order of a single interspike interval, we propose to take single spikes as the basic unit of representation and computations. An alternative would be to consider that probabilities are represented by the average firing rate, defined over long periods of time or large population of neurons (Shadlen & Newsome, 1994). This approach has been fruitful as a description of neural behavior in static situations. Indeed we show that rate coding is a good approximation to neural behavior in our model when the world remains stable for a long period of time. However, we believe it is insufficient to describe online neural computation in an unstable, quickly varying, ambiguous world.

Similarly, we will take single neurons, as opposed to a population of them, as the basic units of computation, considering each neuron as computing the probability of one particular hidden variable. This is in contrast with most alternative approaches that consider populations of neurons as representing probability distributions (Barber, Clark, & Anderson, 2003; Zemel, Dayan, & Pouget, 1998, Sahani & Dayan, 2003; Wu & Amari, 2002) in a cooperative fashion.

We parallel the real, explicit neural space, consisting of neurons, their spike trains and connections, and an implicit probability space, consisting of hidden variables and their statistical dependencies. We propose that the basic meaning of a spike is the occurrence of new, unpredictable probabilistic information and that propagation of spikes in cortical networks corresponds to propagation of beliefs in a corresponding Bayesian network (Frey, 1998; Jordan, 1974; Weiss & Freeman, 2001). We show that this reinterpretation of neural activity and computation provides a new way of looking

at a well-known neural model, the leaky integrate-and-fire neuron, and implies additional nonlinearities that are in accordance with known aspects of cortical physiology.

The model neuron faithfully represents fluctuations in probability of perceptual and motor interpretations of the sensory input. Consequently, this accounts for both the irregularity and apparent noisy nature of neural firing in the presence of static, unambiguous stimuli such as Gabors or optic flow (Vogels, Spilleers, & Orban, 1989; Tolhurst, Movshon, & Dean, 1982), and the much sparser and more precise firing behavior in the presence of noisy, suboptimal, and quickly varying stimuli, such as noisy random dot motion or movies (Bair, 1999; Reinagel & Reid, 2000; Vinje & Gallant, 2002).

In a companion letter in this issue ("Bayesian Spiking Neurons II: Learning"), we show that reinterpreting neural physiology as a form of Bayesian inference allows us to propose a principled form of spike-dependent plasticity where neurons learn to detect patterns of correlations in their synaptic inputs and construct hierarchical causal models for the sensory input over successive neural layers.

2 Bayesian Inference in Single Neurons

We consider that each neuron codes for a time-varying binary hidden variable, x_t. This variable could correspond to a property of the real world, such as the presence or absence of an object in a limited portion of space (the neuron's receptive field) or whether motion goes in one particular direction in the neuron's receptive field. It could also be much more abstract and represent statistical regularities of the sensory input and motor output. Eventually this variable can be learned, in an unsupervised fashion, from the statistics of the synaptic input (see the companion paper). This variable is "hidden" from the neuron that tries to infer its state from its synaptic input. As an illustrative example, we will consider that x_t represents the presence or absence of a horizontal bar at a certain position on the retina.

For clarity, we will distinguish the implicit probability space and its implementation in the real, explicit neural space.

2.1 Implicit Space.

By implicit space, we refer to a quantitative model that describes the statistics of the hidden variable x_t and how it evolves over time, relates to other variables and influences the synaptic input received by the neuron. This model is called generative because it defines the way that observations (the sensory input) are assumed to be generated (or caused) by the state of the hidden variable (Hinton & Ghahramani, 1997). Thus, a generative model might describe how often a horizontal bar appears or disappears at a given retinal location and how its presence result in a particular pattern of light on the retina.

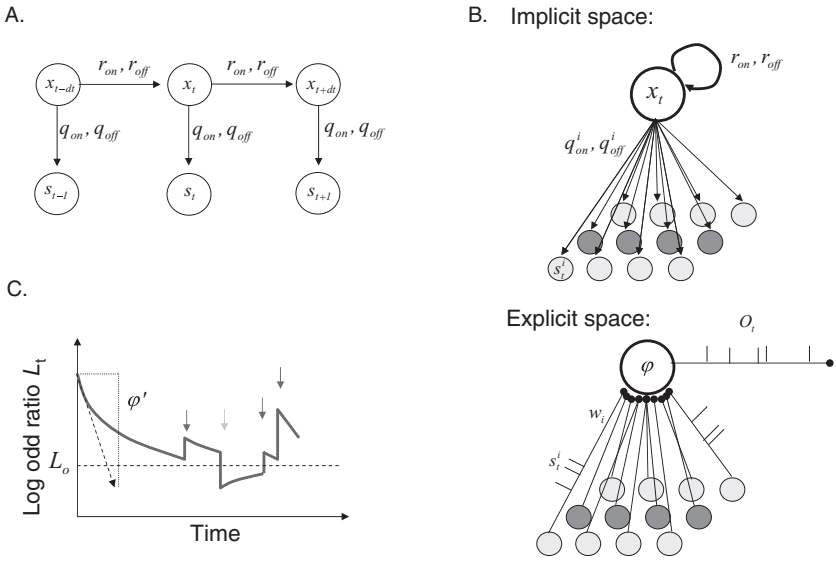

Figure 1: (A) Generative model for the synaptic input received by a single neuron. (B) An illustrative example corresponding to a horizontal bar detector, or a V1 simple cell preferring horizontal orientations. (Top) Generative model in the implicit probability space. $x_t = 1$ corresponds to the presence of a horizontal bar. (Bottom) Spiking neural network implementing Bayesian inference in the explicit neural space. Light gray: inhibitory LGN neuron. Dark gray: Excitatory LGN neurons. (C) Temporal evolution of the log probability ratio. L_t tends to converge toward the prior L_o with a speed defined by φ', depending on the transition rates (see equation 2.2). Dark and light arrows signal the occurrence of excitatory and inhibitory synaptic events.

We assume that the state of this hidden variable at time t, x_t, depends on the state of this variable at the preceding time step, x_{t-dt}, and is conditionally independent of other past states. This is the simplest form of a statistical temporal dependency, corresponding to a Markov chain (see Figure 1A; Hinton & Ghahramani, 1986). A horizontal bar is likely to be present or absent for a certain length of time in the neuron's receptive field. If it is there at time t, it is more likely to be there at time $t + dt$, and vice versa. We call r_{on} and r_{off} the rates at which the state switches on and off, which corresponds to the rate at which bars appear and disappear in our example. In other words, $P(x_t = 1|x_{t-dt} = 0) = r_{\text{on}}dt$ and $P(x_t = 0|x_{t-dt} = 1) = r_{\text{off}}dt$ describe the stability of the hidden variable and are sufficient to account for its temporal statistics.

Second, we consider that the state of the hidden variable, which cannot be observed directly by cortical neurons, causes (i.e., results in) a particular sensory input, which, for a neuron, takes the form of synaptic events (spikes)

received from a collection of N synapses. In our toy example, the presence of a horizontal bar implies a particular distribution of local contrasts on the retina and, thus, particular synapses coming from lateral goniculate nucleus (LGN) neurons to receive action potentials at higher or lower rates.

We represent this synaptic input by a vector of binary variable $\mathbf{s}_t = [s_t^i]_{i=1,\ldots,N}$, where $s_t^i = 1$ when the synapse number i is activated during time t and $t + dt$. Each synapse is activated with a particular probability, $P(s_t^i = 1|x_t = 1) = q_{\text{on}}^i dt$ if the state is 1 and $P(s_t^i = 1|x_t = 0) = q_{\text{off}}^i dt$ if the state is 0. Since there is no direct temporal dependency between $\mathbf{s}_t$ and $\mathbf{s}_{t+dt}$, the synaptic input can be described as an inhomogeneous Poisson process with rates q_{on}^i when the state is 1 and q_{off}^i when the state is 0. The corresponding model is a hidden Markov chain (see Figure 1A), which describes how the synaptic input was generated. This is the generative model of the sensory input, $\mathbf{s}_t$.

For example, we might consider that our horizontal bar–specific neuron receives inputs from a set of center-surround LGN neurons, as in the classical Hubel and Wiesel (1970) model of a simple cell (see Figure 1B). The synapses from LGN neurons in the center of the simple cell's receptive field (dark gray in Figure 1B) are more active when a bar is present in the center of the receptive field, that is, when $x_t = 1$, which corresponds to the fact that $q_{\text{on}}^i > q_{\text{off}}^i$ for these neurons. On the other hand, neurons with receptive fields at the periphery (light gray neurons in Figure 1B) are less active when the central bar is present—$q_{\text{on}}^i < q_{\text{off}}^i$.

Inference in this hidden Markov model can be performed by a recurrent process. In particular, we can compute the log-odds ratio of the hidden state at time t, L_t, given all synaptic inputs received in the past. L_t is defined as

$$L_t = \log\left(\frac{P(x_t = 1|\mathbf{s}_{0\to t})}{P(x_t = 0|\mathbf{s}_{0\to t})}\right), \tag{2.1}$$

where $s_{0\to t}$ corresponds to the synaptic input received from time 0 to time t. If we take the limit of the temporal update equations as $dt \to 0$, we get the following differential equation (see appendix A):

$$\begin{aligned}\dot{L} &= r_{\text{on}}(1 + e^{-L}) - r_{\text{off}}(1 + e^{L}) + \sum_i w_i \delta\left(s_t^i - 1\right) - \theta \\ &= -\varphi(L_t) + I_t. \end{aligned} \tag{2.2}$$

w_i, the synaptic weights, describe how informative a synapse i is about the state of the hidden variable, for example, $w_i = \log(\frac{q_{\text{on}}^i}{q_{\text{off}}^i})$. Each synaptic input ($s_t^i = 1$) gives an impulse to the log-odds ratio, which is positive if this synapse is more active when $x_t = 1$, that is, when $q_{\text{on}}^i > q_{\text{off}}^i$ (as it increases the neuron's confidence that the state is 1). On the contrary, if the impulse is negative, the neuron's confidence is decreased if this synapse is more

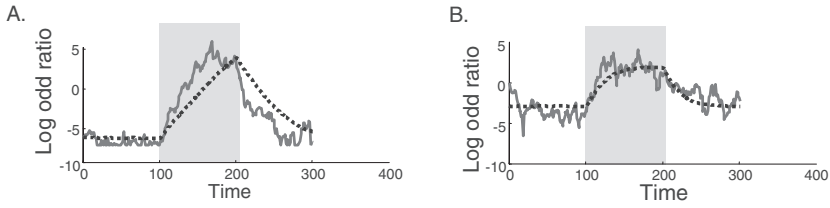

Figure 2: The log-odds ratio reflects a leaky integration of sensory evidence. (A) An example with small transition rates and long effective integration time constant ($r_{\text{on}} = 0.00005, r_{\text{off}} = 0.0001$). (B) An example with larger transition rates and shorter effective time constant ($r_{\text{on}} = 0.01$, $r_{\text{off}} = 0.02$). Solid line: log-odds ratio on a single trial. Dotted line: log-odds ratio averaged over 1000 trials.

active when $x_t = 0$, that is, when $q^i_{\text{on}} < q^i_{\text{off}}$. Thus, synaptic inputs from the light gray LGN neurons to the simple cell in the figure are inhibitory, with negative weights, while inputs from the dark gray LGN neurons are excitatory.

The term φ depends on the transition rates and implements a temporal "leak" for L_t. The bias, θ, is determined by how informative it is not to receive any spike, $\theta = \sum_i q^i_{\text{on}} - q^i_{\text{off}}$. The synaptic drive, $I_t = \sum_i w_i \delta(s^i_t - 1) - \theta$, is the total contribution of sensory observations.

2.2 Explicit, Neural Space. We propose that neural dynamics in the explicit neural space implements the inference and learning in the underlying generative model described in Figure 1A. Thus, equation 2.2 corresponds to a leaky synaptic integration process: the first part of the equation ($\varphi(L_t)$) depends on the temporal statistics of the hidden state (r_{on}, r_{off}) and on the log probability ratio itself, L_t. The overall effect of this component is to bring L_t back toward a prior level $L_o = \log(\frac{r_{\text{on}}}{r_{\text{off}}})$, which corresponds to what is known about the hidden state in the absence of any observation (see Figure 1C). This results in a gradual forgetting or fading of the neuron's certainty about its hidden state, which is faster when the world changes rapidly, that is, when the transition rates are high. The second part of the equation corresponds to the synaptic drive and is a weighted sum of the contribution from all synapses. Its overall effect is to drive L_t away from its prior.

In order to test the neural response in conditions similar to neurophysiological studies, the state of x_t was fixed at 1 when a stimulus was presented in the neuron's receptive field at time t and 0 otherwise (thus, the temporal profile of x_t is defined by the experimental protocol). The synaptic input $\mathbf{s}_t$ was sampled from Poisson processes with rates q^i_{on} when $x_t = 1$ and q^i_{off} when $x_t = 0$. The temporal profile of L_t during presentation of the preferred stimulus is plotted in Figure 2.

On a single trial, L_t fluctuates over time to reflect random arrivals of excitatory and inhibitory synaptic inputs (see Figures 2A and 2B). On average, though, the log-posterior ratio reflects the leaky integration of synaptic evidence, with an effective time constant that depends on the transition probabilities $r_{\text{on}}, r_{\text{off}}$. If the state of x_t is very stable ($r_{\text{off}} \sim 0$), synaptic evidence is integrated over almost infinite time periods, the mean log posterior ratio tending to increase or decrease linearly with time (see Figure 2A). In the example in Figure 2B, however, the state is less stable, so "old" synaptic evidence is discounted and L_t saturates.

This is reminiscent of the leaky integration of synaptic inputs in biological neurons. However, to understand the neural basis of probabilistic computation, we need to define the rules according to which a neuron will fire output spikes as a function of its synaptic inputs, that is, what relates its output spike train O_t with the synaptic input $\mathbf{s_t}$. In other words, what is the neural code? This output spike train should provide a good representation of L_t, since it is all that will be available for performing further probabilistic computations.

2.2.1 Spike Generation. We use the same convention for O_t as for s_t^i, for example, $O_t = 1$ when an output spike is fired between time t and $t + dt$, and 0 otherwise.

Predictive coding. We propose that each spike deterministically reports new information about the state x_t that is not redundant with what was already reported by the preceding spikes. In other words, the neuron performs a form of predictive coding and fires only when it cannot predict itself. This corresponds more or less to having spikes represent the temporal derivative of L_t.

Intuitively, this ensures that the model is self-consistent, in the sense that the output of the Bayesian neuron can be used as an input for another Bayesian neuron. If spikes represent only new information rather than an integration of sensory evidence, they can be harmlessly integrated in later processing stages without running into the problem of redundant successive integrations. The cost of this neural code is that exact inference can be performed only in a limited family of generative models, where only the objects highest in the hierarchy truly have a temporal dynamics (see Figure 6B).

In order to fire only when new information is available, we propose that a neuron implements a form of spike-dependent adaptation, increasing its firing threshold after each spike. Thus, the neuron compares online the odds for its hidden variable, L_t, with a prediction G_t computed from the output spike train. A spike is emitted when the odds (a leaky integration of the synaptic input s_t) exceeds the prediction (a leaky integration of the output spike train O_t). We defined the prediction G_t as what another neuron would obtain as estimates for L_t if it was integrating the output spikes, with a synaptic weight of g_o and bias of 0 (as in equation 2.1, but applied

to the output spike train rather than the synaptic inputs). In this way, our neuron fires only when what it "believes" (L_t) would not match what another neuron could learn from its previous spikes (G_t). This reasoning is illustrated in Figure 3A.

Thus, the dynamical equations relating the input and the output spike trains are:

$$\begin{aligned}
\dot{L} &= r_{\text{on}}(1+e^{-L}) - r_{\text{off}}(1+e^{L}) + \sum_i w_i \delta\left(s_t^i - 1\right) - \theta \\
\dot{G} &= r_{\text{on}}(1+e^{-G}) - r_{\text{off}}(1+e^{G}) + g_o \delta(O_t - 1) \\
O_t &= 1 \text{ if and only if } L_t > G_t + \frac{g_o}{2}.
\end{aligned} \tag{2.3}$$

Here g_o, a positive constant, is the only free parameter, the other parameters being constrained by the statistics of the synaptic input s_t.

Figure 3B plots a typical trial, showing the behavior of L, G, and O before, during, and after the presentation of the stimulus. As random synaptic inputs are integrated, L fluctuates and eventually exceeds G, leading to an output spike. Immediately after a spike, G jumps to $G + g_o$, which prevents (except in very rare cases) a second spike from immediately following the first. Thus, this "jump" implements a relative refractory period, or a spike frequency adaptation mechanism. However, G decays as it tends to converge back to its stable level (the prior) $g_s = \log(\frac{r_{\text{on}}}{r_{\text{off}}})$. If new positive synaptic evidence arrives in the form of spikes at excitatory synapses, L eventually exceeds G again, leading to a new spike. This threshold crossing happens more often during stimulation ($x_t = 1$) as the net synaptic input alters to create a higher overall level of certainty, L_t. As illustrated in Figure 3B, the prediction G_t tracks the log odds and provides a rough approximation when the log odds is above the prior L_o.

Similar mechanisms have been proposed to describe the effects of relative refractory periods (Gerstner & Werner, 2002) and spike-dependent adaptation (Liu & Wang, 2001). Our approach proposes a specific computational role for these forms of adaptation. This firing mechanisms allow Bayesian neurons to be self-consistent: another neuron, integrating the output spike train O_t with equation 2.2, would recover an estimate of G_t and thus an estimate of the log odds L_t, since G tracks L. The same operations can be applied on the input and the output spike trains, which would not be the case if the neuron's firing rate represented L_t explicitly (see section 3).

In the companion letter in this issue, we show that the log odds can be decoded even more accurately, including log odds that are below the prior $L_o = \log(\frac{r_{\text{on}}}{r_{\text{off}}})$. Note also that an efferent neuron needs to know r_{on} and r_{off} to decode the output spike train O_t. This information is not directly available in the output spike train but can be learned by the efferent neuron (see the companion letter).

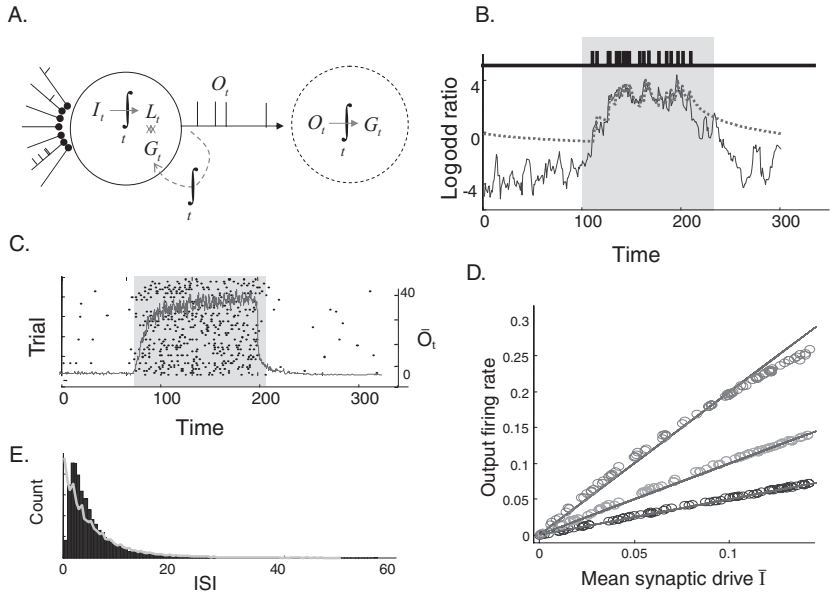

Figure 3: Predictive coding: Mechanism and prediction for firing statistics. (A) Diagram illustrating the principle of predictive coding. G_t is both what another neuron would know about the hidden variable x_t by observing the output spikes and the firing threshold. (B) An example trial. Solid line: L_t. Dotted line: G_t. Thick black line: output spikes fired on this trial. Shaded area indicates periods when $x_t = 1$. Parameters were $x_t, d_t = 0.1, r_{\text{on}} = 0.001, r_{\text{off}} = 0.01, g_o = 1.5$, $N = 50, q^i_{\text{on}} = 0.06$, and $q^i_{\text{off}} = 0.03$. (C) An example output spike raster plot (black dots represent spikes, one trial on each line) and mean firing rate (solid line) of the neuron. The mean input firing rate is obtained by averaging the number of spikes fired in temporal window $dt = 0.1$ ms for 10,000 trials (D) Mean output firing rate (number of spikes fired during one second, averaged over 1000 trials) during presentation of the stimulus ($x_t = 1$) as a function of the mean synaptic drive for a random selection of parameters and for three different g_o. Top circles and line: $g_o = 0.5$; middle circles and line: $g_o = 1$; bottom circles and line: $g_o = 2$. The lines represent the prediction of the model $\bar{O} = \frac{\bar{I}}{g_o}$. Each data point represents a random selection of parameter: r_{on} and r_{off} between 0.1 and 0.00001; $N = 20$; $dt = 0.1$; q^i_{off} between 0.005 and 0.05 for each i. $q^i_{\text{on}} = \exp(w_i) q^i_{\text{off}}$ where w_i is randomly selected between -0.5 and 0.5. Because of sampling problems, sets of parameters leading to firing rates bigger than $\frac{0.1}{dt}$ were excluded. (E) Interspike interval distribution for the output spikes, computed during presentation of the preferred stimulus (parameters as in Figure 3C). The light line indicates the ISI distribution of a Poisson process with the same firing rate.

The efficiency of this encoding depends directly on the variable g_o, which defines how many spikes a neuron is willing to fire in order to represent the odds. If g_o is very small, the decoding is more precise, but the encoding is expensive, given that there are potentially as many or more output spikes in O_t than there were input events in s_t. If g_o is large, the representation is less accurate, since small fluctuations in log odds can lead to big jumps in the prediction. However, there are many fewer output spikes than input spikes, that is, the information is greatly compressed, for a very limited information loss.

For the simulations presented in this letter, we chose $g_o = 1.5$ (except otherwise stated). Thus, whenever neurons fire a spike, its predicted odds G_t increases by 1.5. In other words, an output spike expresses a multiplication of the odds ($\frac{P(x_t=1)}{P(x_t=0)}$) by $e^{1.5}$.

2.2.2 Prediction for the Output Firing Statistics

Mean firing rate. The instantaneous firing rate $\bar{O}_t$ of the Bayesian neuron during presentation of its preferred stimulus (i.e., when x_t switches from 0 to 1 and back to 0) is plotted in Figures 3C and 3D.

In contrast with the log-odds ratio (see Figure 2), the mean output firing rate $\bar{O}_t$ tracks the state of x_t almost perfectly rather than exhibiting slow rise and fall when the hidden state switches on or off (see Figure 3C). This is because, as a form of predictive coding, the output spikes reflect the new synaptic evidence, contained in the total input $I_t = \sum_i w_i \delta(s_t^i - 1) - \theta$, rather than the log-odds ratio itself.

Moreover, the mean output firing rate is very close to a rectified linear function of the mean input, $\bar{O} = \frac{1}{g_o}[\bar{I}]^+ = \frac{1}{g_o}[\sum_i w_i q^i_{\text{on(off)}} - \theta]^+$. This is illustrated in Figure 3D for random selections of parameters and three different g_o. This relationship between $\bar{O}$ and the mean level of evidence $\bar{I}$ received by the neuron is a direct consequence of the fact that G tracks L and follows the same dynamics (see section 3 and appendix B).

Note that when $x_t = 1$, the mean synaptic input is given by (before taking the limit for small dt; see appendix A)

$$\bar{I}^{\text{on}} = \frac{1}{dt} \sum_i q^i_{\text{on}} dt \log\left(\frac{q^i_{\text{on}}}{q^i_{\text{off}}}\right) + \left(1 - q^i_{\text{on}} dt\right) \log\left(\frac{1 - q^i_{\text{on}} dt}{1 - q^i_{\text{off}} dt}\right). \tag{2.4}$$

From this we can conclude that the mean input when $x_t = 1$ is also the Kullback-Leibler (KL) divergence between the probability of the instantaneous synaptic input when the state is 1 and 0, that is,

$$\bar{I}^{\text{on}} dt = \text{KL}(p(\mathbf{s}_t | x_t = 1) | p(\mathbf{s}_t | x_t = 0)), \tag{2.5}$$

and, vice versa, when $x_t = 0, \bar{I}^{\text{off}} dt = -\text{KL}(p(\mathbf{s}_t | x_t = 0) | p(\mathbf{s}_t | x_t = 1))$.

Our model predicts neurons that respond as long as they receive evidence that their preferred stimulus is present and stop responding when their stimulus disappears, with rather sharp transitions rather than slow ramps between active and inactive states. This is easy to understand intuitively: as long as the stimulus is present, the new sensory inputs received since the last output spike result in an increase of the log odds L_t. Meanwhile, the predicted log odds G_t decreases as the neuron forgets the contribution of its last output spike. This will lead to a new output spike when L_t crosses G_t once again.

Note that this model neuron signals unpredictable changes of the probability of $x_t = 1$, not unpredictable changes in the state x_t itself. A neuron using predictive coding for the state rather than the probability of x_t would signal changes in this state, for example, switches from $x_t = 0$ to $x_t = 1$. Thus, this neuron would fire only at the time when the stimulus appears and become silent during longer presentation of the stimulus. A large class of sensory neurons may match this description, and we intend to explore this alternative hypothesis in future work.

This is in contrast with neurons whose firing rate is proportional to L_t. The firing rate of such neurons would increase linearly when their stimulus is present and decrease linearly when their stimulus is absent (but eventually saturate). There is evidence for this last form of coding in sensorimotor areas during slow integration tasks (Mazurek, Roitman, Ditterich, & Shadlen, 2003). However, in addition to the computational problems posed by this kind of representation (see section 3), it seems that sensory cortical neurons are more likely to fit the first description than the second.

Poisson-like statistics. While the firing rate of the Bayesian neuron can simply be described as a function of the hidden state, the structure of the spike trains themselves looks very irregular and unpredictable from trial to trial, as illustrated in Figures 3C and 3E. During periods when the preferred stimulus is present (stimulus-driven response) or absent (rest), the neuron's firing appear to be memoryless and close to a Poisson process. In particular, we found a Fano factor close to 1 and a quasi-exponential interspike interval (ISI) distribution (see Figures 3D and 3F) for a wide range of parameters. This prediction can appear surprising given that the firing time is a deterministic function of the synaptic input. However, this unpredictable firing is a direct consequence of the unpredictable arrival of Poisson-distributed synaptic events, that is, of the input noise. Insights into why this is the case can be obtained by comparing the Bayesian neuron with a classical neural model, the leaky integrate-and-fire neuron.

2.3 Similarity with Leaky Integrate-and-Fire Neurons. In this section we show the similarities of the Bayesian neuron with a leaky integrate-and-fire (LIF) neuron. LIF neurons integrate their synaptic input linearly,

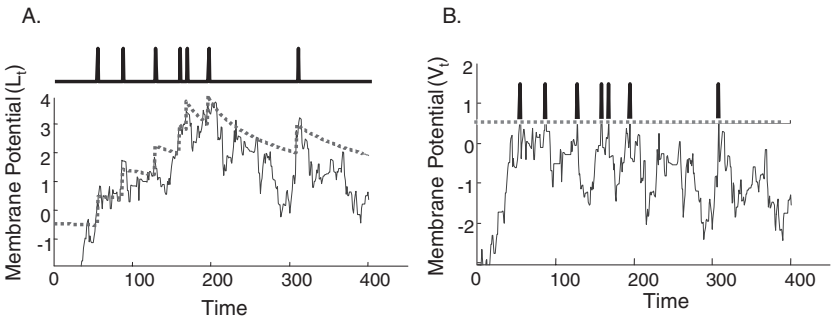

Figure 4: Equivalence with leaky integrate-and-fire neuron: an example trial. (A) Interpretation as an adapting integrate-and-fire neuron with time-varying threshold. Solid line: membrane potential (Lt). Dotted line: threshold (Gt). Thick solid lines: spikes (O_t). (B) Same neuron and same trial in A. Interpretation as a leaky integrate-and-fire neuron with a reset. Solid line: membrane potential (V_t). Dotted line: Constant threshold. Solid thick lines: Spikes O_t. The parameters in this example are $r_{\text{on}} = 0.003$, $r_{\text{off}} = 0.005$, $N = 100$, $q^i_{\text{on}} = 0.4 + \text{sign}(i - 70) * 0.1$, $q^i_{\text{off}} = 0.4 - \text{sign}(i - 70) * 0.1$, and $x_t = 1$.

with a particular time constant τ, so that the temporal evolution of their membrane potential between two spikes is expressed by

$$\tau \frac{\delta V}{\delta t} = -(V - V_{\text{rest}}) + I_n(t), \tag{2.6}$$

where $I_n(t)$ is the average current resulting from presynaptic spiking. When V reaches a particular threshold, a spike is fired and V is reset to a reset potential V_{reset}.

Our goal is not to link arbitrarily different probabilistic messages with their biophysical counterparts, such as the membrane potential. Actually the integrate-and-fire neuron, while being useful for its mathematical tractability, is far from describing the dynamics of real neurons (Aguera y Arcas, Fairhall, & Bialek, 2003). Rather, we would like to use the analogy for making further predictions, explain why the firing of Bayesian neurons is close to memoryless, and express deviation from linearity predicted by the model.

As a first approximation, we could interpret L_t as the membrane potential of the neuron and G_t as a time-varying threshold for spike generation, increasing after each spike (see Figure 4A). Threshold increasing after each spike has been proposed to implement both refractory periods (Gerstner & Werner, 2002) and spike-dependent adaptation (Liu & Wang, 2001). Interestingly, our model suggests that such a mechanism might implement predictive coding and ensure that the output spike trains are properly decoded by

efferent neurons. Note that in contrast with previous models, the Bayesian neuron has no reset, that is, the membrane potential does not go back to a rest potential after each spike.

However, we can gain further insights into the computation performed by this neuron by using a completely different interpretation of the membrane potential. Let us consider that the membrane potential is not L directly but the difference between L and the prediction G. In other words, the membrane potential of the neuron is the difference between what the neuron believes about the hidden state (the current log odds that needs to be conveyed to other neurons) and what the neurons have already told using its previous spikes (the prediction G). This implies a certain temporal independence between successive interspike intervals.

Here we reduced the analysis to prolonged, statistically stable periods when the state is ON ($x_t = 1$). We define the neutrally stable certainty, $\bar{L}$, as the value of L for which the derivative $\dot{L}$ is zero on average. When $\bar{L} \approx \bar{G}$, g_o and the input fluctuations can be considered small compared to $\bar{L}$, we can approximate the neural dynamics by the following equation for the membrane potential $V_t = L_t - G_t$ (see appendix B),

$$\begin{aligned} &\tau_{\bar{L}} \dot{V}_t = -V_t + \tau_{\bar{L}} I_t \\ &V_t > \frac{g_o}{2} \Rightarrow V_t = -\frac{g_o}{2} \text{ and } O_t = 1, \end{aligned} \tag{2.7}$$

where $\tau_{\bar{L}}$ is the temporal constant of the membrane potential and I_t is the synaptic drive to the neuron, as defined previously. Note that I_t is a weighted sum of synaptic input minus the bias θ. From this equation, we can also conclude that since $\bar{V} = \bar{L} - \bar{G} \approx 0$, we have $\bar{O} \approx \frac{\bar{I}}{g_o}$, in agreement with simulation results (see Figure 3E).

Additional insights can be obtained if two other conditions are met (see appendix B): $g_o \approx 2\frac{e^{\bar{L}}+1}{e^{\bar{L}}}$ and the probability of x_t being 1 is relatively high, for example, $\bar{L} > L_o + 1$. We can then approximate the neural dynamics by the following equation (see appendix B),

$$\begin{aligned} &\tau_{\bar{L}} \dot{V} = -\left(V - \frac{g_o}{2}\right) + \tau_{\bar{L}}(I_t - \bar{I}) \\ &V_t > \frac{g_o}{2} \Rightarrow V_t = -\frac{g_o}{2} \text{ and } O_t = 1 \end{aligned} \tag{2.8}$$

where $\bar{I}$ is the mean synaptic drive. The "rest" potential $V_{\text{rest}} = \frac{g_o}{2}$ is also the the spike threshold.

This neuron is an integrate-and-fire neuron: it integrates the synaptic input linearly with a temporal constant $\tau_{\bar{L}}$, fires a spike when this membrane potential reaches a particular threshold ($\frac{g_o}{2}$ in this case), at which time the membrane potential is reset to $-\frac{g_o}{2}$ (see Figure 4B). Note, however, that

two characteristics distinguish this neuron's computation from a classical integrate-and-fire neuron.

The first characteristic is an input-dependent time constant. The temporal time constant of integration, $\tau_{\bar{L}}$, depends on the transition rates, r_{on} and r_{off}, but also on the mean level of certainty L_t. More precisely, it is given by

$$\tau_{\bar{L}} = \frac{1}{r_{\text{on}} e^{-\bar{L}} + r_{\text{off}} e^{\bar{L}}}. \tag{2.9}$$

If the neuron receives stronger excitation than inhibition, that is, if L_t is strongly positive, or if it receives stronger inhibition than excitation, that is, if L_t is strongly negative, the neuron acts as a coincidence detector and changes its level of certainty only when receiving several coincident inputs of the same valence (either several excitatory inputs or several inhibitory inputs).

If, on the other hand, the neuron's log-odds ratio is small in terms of absolute value, that is, if the neuron receives approximately the same amount of positive or negative evidence, the time constant is longer. The neuron acts like an integrator when it is uncertain of the state of the hidden variable, taking into account each new piece of evidence that comes in. The temporal constant is minimal for $L = \frac{L_o}{2}$.

Interestingly, a time constant that depends of the level of activation is a characteristic that distinguishes real neurons from the classical integrate-and-fire model. When receiving strong excitatory or inhibitory currents, biological neurons have a conductance that goes up (due to the openings of ion channels) and thus a time constant that goes down. Cases when $\bar{L}$ is strongly positive or negative, and thus the time constant is short, correspond to cases where neurons receive strongly dominant excitation or strongly dominant inhibition.

A second interesting characteristic of our model neuron is that in stable and highly informative regimes, when it can be described by equation 2.8, it is driven by the synaptic inputs minus the mean input, that is, $I_t - \bar{I}$ (see appendix B). A neuron is thus caused to spike not by the progressive integration of synaptic spikes it receives but by fluctuation of this synaptic drive around its mean, that is, a balanced input. In particular, the membrane potential follows a random walk around the spike threshold.

Thus, we can predict that neural firing will be driven by input fluctuations and close to memoryless, with the Fano factor around 1, as in the case of integrate-and-fire neurons driven by noise (Shadlen & Newsome, 1994). This is in contrast with classical integrate-and-fire neurons whose firing is regular when uncorrelated excitatory input dominates (Zohary, Shadlen, & Newsome, 1994).

It is easy to understand why it should be so. The model neuron implements a form of predictive coding and fires only when the integrated input exceeds a prediction from the past output spikes. Thus, in effect, it

is canceling out from its current input the mean input received in the past. Only unpredictable fluctuations lead to an output spike.

It has been proposed that the Poisson-like firing statistics observed in the cortex result from neurons with balanced excitation and inhibition around their threshold (Shadlen & Newsome, 1994). If excitation and inhibition were not balanced around the firing threshold, neurons would behave as integrators and generate much more regular spike trains. Indeed, a tight balance between excitatory and inhibitory conductances has been found experimentally in neurons (Tao & Poo, 2005) and even in individual dendritic branches (Liu, 2004). We propose a different, albeit closely related, idea, where neurons will fire with Poisson statistics even if they do not receive balanced synaptic input. However, because they use a form of predictive coding and inhibit themselves through the adaptive mean input $\bar{I}$, they will respond to fluctuations of their input in the same fashion as balanced integrate-and-fire neurons.

Note that while the Bayesian neuron balances itself in a statistically stable regime, it is not the case in nonstable regimes, when x_t rapidly switches state. Thus, when x_t switches ON, the neuron suddenly receives a strong wave of excitation. This will result in a sharp rise of the log odds, tracked by the prediction with a delay. The spikes in response to this transient will be more delayed and more temporally precise than expected if the input was perfectly balanced. In response to a quickly varying state, the behavior of the neuron will significantly deviate from Poisson (see Figure 5).

2.4 Similarity with Rate Coding. The model neuron's output firing statistics, in response to a Poisson distributed input, is close to a Poisson process (see Figures 3C and 3E). The mean output firing rate depends on the hidden state and is conditionally independent of time. Moreover, the firing rate can be described as a linear rectified function of the mean synaptic drive or mean rate of evidence received by the neuron, $\bar{I}$. In this condition, one might wonder if it is really necessary to consider individual spikes: the output is not qualitatively different from a rate code model, where the firing rate provides information about x_t.

However, obtaining this input-output function is a nonlinear transform that requires L_t and G_t as intermediate stages. In effect, it consists of selecting among a bombardment of weakly informative synaptic input exactly what is relevant for the hidden variable. This selective evidence is expressed in the output spike train, with typically many fewer spikes than in the synaptic input (this is quite important, given that cortical neurons receive hundreds of active connections and can fire only a few tens of spikes in a second). This computation conserves information rather than adds noise by sampling spikes at random from a particular rate.

More importantly, the Poisson statistics of the output spike train is a direct consequence of the noise in the synaptic input. If, rather than changing from trial to trial, the synaptic input was identical (frozen noise), the

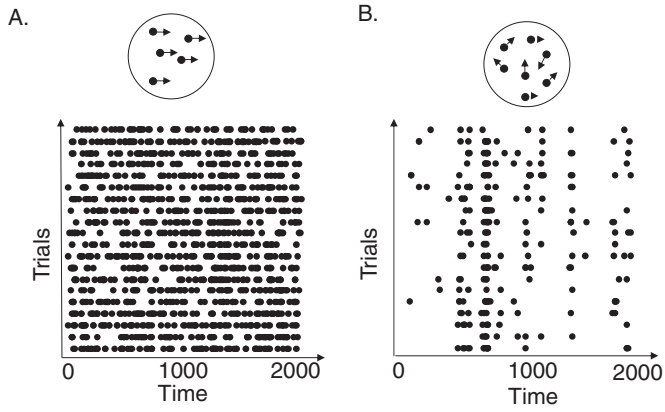

Figure 5: Interspike intervals faithfully reflect fluctuations in probability. (A) Spike raster plot of a model MT neuron receiving constant motion stimulus in its preferred direction. (B) Spike raster plot of a model MT neuron receiving random dot motion stimulus with frozen noise. In both cases, the neuron receives $N = 100$ synapses—two for each possible 50-dot location. At each dot location, one synapse is excitatory ($w_i = 0.5$) and fires at 30 spikes per seconds ($q^{\mathrm{i}}_{\mathrm{on}} = 0.03$) when a dot moves in the neuron's preferred direction, and 18 spikes per seconds otherwise. Another synapse is inhibitory ($w_i = -0.5$) and fires with 30 spikes per seconds when a dot moves in the neuron's antipreferred direction and 18 spikes per seconds otherwise. The transition rates of the MT neuron are set to $r_{\mathrm{on}} = 0.005$, $r_{\mathrm{off}} = 0.01$. In the constant motion condition, all the dots are moving in the neuron's preferred direction. In the random dot motion condition, half of the dots (25) are moving in the preferred direction and the other half in the antipreferred direction. Dots appear independently at each location with probability $p = 0.006$ and stay on the screen for 100 ms. Time axis is in ms.

output spikes would occur at exactly the same time. This also means that fluctuations in interspike intervals are meaningful in our model: they carry information about the precise temporal structure of the sensory input, corresponding to fluctuations of probability for the hidden variable.

As a consequence, the same model that predicts Poisson-distributed spikes for noiseless stimuli predicts a more reproducible firing response to noisy stimuli. It is illustrated as a toy example in Figures 5A and 5B. This example reproduces an interesting result obtained by Bair (1999). Here an example mediotemporal (MT) area neuron codes for a hidden variable corresponding to the presence or rightward motion in the scene, the stimulus corresponding to a pattern of moving dots. For simplicity, we assumed that the MT neuron receives inputs from a set of local V1 motion detector. These V1 cells fire spikes according to an inhomogeneous Poisson process, with a rate that depends on the presence in their receptive field of a dot moving in

their preferred direction. As expected, the response of a rightward motion-selective cell to constant rightward optic flow looks Poisson, fitting the predictions of a rate model. In this case, the probability of rightward motion is constant and high, and fluctuations in this probability are due to synaptic inputs arriving at unpredictable times from V1 motion detectors. On the other hand, the response of the same cell to a repeated random dot motion stimulus appears much more reproducible. The spikes signal periods of increase in the probability of the preferred motion due to spurious correlations between the moving dots. These fluctuations in the stimulus are relevant if the animal is trying to detect subtle and short events.

3 Discussion and Conclusion

We started from an interpretation of synaptic integration in single neurons as a form of inference in a hidden Markov chain. We derived a model of spiking neurons able to compute the marginal posterior probabilities of sensory and motor variables given evidence received in the entire network. In this view, the brain implements an underlying Bayesian network in a neural architecture, with conditional probabilities represented by synaptic weights. The model makes a rich set of predictions for the general properties of neuron and synaptic dynamics, such as a time constant that depends on the overall level of inputs, specific forms of frequency-dependent spike and synaptic adaptation (not shown here), and balanced excitation and inhibition. However, it is still restricted to probabilistic computations involving binary variables. In a related work, similar ideas are applied to population encoding of log probability distribution for analog variables (Huys, Zemel, Natarajan, & Dayan, 2007).

Despite nonlinear processing at the single cell level, the emerging picture is relatively simple: the neuron acts as a leaky integrate-and-fire neuron driven by noise. The output firing rate is a rectified weighted sum of the input firing rates, and the firing statistics are Poisson. However, these output spike trains are a deterministic function of the input spike trains. Spikes report fluctuations in the level of certainty that could not be predicted from the stability of its stimulus (contribution from G_t). Thus, firing will be, by definition, unpredictable. This last observation leads us to suggest that the irregular firing and Poisson statistics observed in cortical neurons (Britten, Shadlen, Newsome, & Movshon, 1992) arise as a direct consequence of the random fluctuations in the sensory inputs and the instability of the real word, but are not due to unreliable or chaotic neural processing.

3.1 Neural Representation of Probability. What is the neural representation of probabilities? We propose that the probabilities of perceptual or motor variables are not represented explicitly in the output firing rates of the neurons. Rather, they correspond to an internal activation level of the neuron, which is not directly observable except by integrating its output spike

train. The parameters of this integration, and thus what a spike means, are learned online by efferent neurons. Thus, we propose that neurons and neural networks are highly adaptive structures that continuously change their dynamical properties in order to interpret their input as best as possible.

Our model neurons are responding as long as they receive evidence in favor of their hypothesis, with a firing rate proportional to the strength of the evidence. In contrast, previous models had proposed that firing rates explicitly represent probabilities (Mazurek et al., 2003; Rao, 2003; Zemel et al., 1998; Sahani & Dayan, 2003). This predicts firing rates that increase during stimulus presentation to reflect the accumulation of evidence. Such neurons have been reported in experiments with monkeys trained to perform slow-motion integration tasks (Mazurek et al., 2003; Shadlen & Newsome, 2001). In these experiments, the animals were required to signal with an eye movement the perceived direction of motion of a noisy display. It was found that cells in the lateral intraparietal cortex (LIP), an area linked to the planning of eye movements, had firing rates that increased over the duration of motion integration, with a slope proportional to the strength of the evidence (i.e., the motion coherence in the display). Thus, these cells acted as integrators, whose response could be interpreted as a log probability ratio. In contrast, cells in the MT had a firing rate that reflected the motion coherence but did not increase or decrease during integration.

The Bayesian neuron described here would be a better description of the MT cells rather than the LIP cells. In fact, the first type of responses is a general feature of visual areas, while integrator-like responses are a feature of sensorimotor area involved in triggering the behavioral output. As we argue in the next paragraph, the integration stage needs to be reserved for the last stage in a hierarchy of inference in time. Eventually the probabilities will need to become explicit in order to be translated into behavioral choices. However, we propose that they should not be explicit in the early stages of sensory processing, even if all cortical neurons are involved in Bayesian computations.

3.2 Why Not a Rate-Based Model? One alternative neural coding of probability could be to fire spikes stochastically, with a probability that is proportional to (or a function of) the log probability ratio L_t. Such a rate-based rule has been proposed by Rao (2003) in a related model. More generally, rate coding is, to our knowledge, the only form of probabilistic encoding that has been considered so far (Zemel, Dayan, & Pouget, 1997; Barber et al., 2003; Sahani & Dayan, 2003). It is compatible with some neurophysiological data, since firing rates in the lateral parietal areas have been reported to be proportional to the log probability ratio for appropriate eye movement responses (Mazurek et al., 2003) or probability of reward (Glimcher & Rustichini, 2004; Sugrue, Corrado, & Newsome, 2004).

This encoding is seductive in its simplicity. However, it has two major drawbacks. First, being a stochastic spike generation rule, it adds

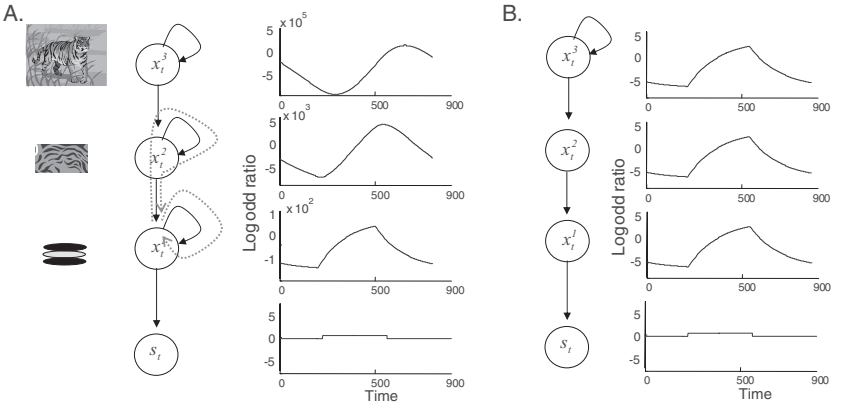

Figure 6: Why rate coding is not a good solution. (A) Toy example where a horizontal bar can be caused by stripes, which themselves can be caused by a tiger. We illustrate here the consequence of transmitting directly L_t, in the form of a firing rate, from neuron to neuron. Dotted arrows: two redundant paths of integration. See the text. (B) The same toy example when predictive coding is used. Here the temporal derivative of L_t is transmitted from neuron to neuron, and the subsequent integration recovers L_t.

uncertainty, and thus noise, to an otherwise deterministic probability computation. Second, and more importantly, the resulting model would not be self-consistent since the input and output firing rates have different meanings and different dynamics. The input spike rates q^i_{on},q^i_{off} are constant and a function of the state. The output firing rate is a leaky integration, growing and decaying over time to reflect an accumulation of evidence about the state (see the dotted line in Figure 2).

Figure 6A is a toy example showing the consequence of this inconsistency. The output of the horizontal bar neuron is sent to a neuron coding for the presence of horizontal stripes. In turn, the output of the stripe neuron is sent to a neuron coding for the presence of a tiger. If we suppose that the output firing rate from the horizontal bar neuron is proportional to the log probability ratio, we can describe the input-output relationship of the neuron as a leaky integration. Thus, the output firing rate of the horizontal bar neuron will consist of a linear ramp, starting at the presentation of the preferred stimulus, followed by a saturating plateau. Integrator neurons of this type have been reported (Mazurek et al., 2003). However, the firing rate of the stripe neuron reflects a double integration of sensory evidence, since it integrates the output of the horizontal bar neuron and grows quadratically with time. The tiger neuron reflects a triple integration, and so on. As a result, the firing rates grow far too much compared to the actual information contained in the sensory input, while the delay in response due to the slow rise of integration becomes increasingly long.

This problem has its corresponding interpretation in the implicit probability space. In all logic, a tiger "causes" the presence of stripes, which themselves cause the presence of horizontal bars. The corresponding generative model is a network of a coupled hidden Markov chain (see Figure 6A). The "naive" form of inference by successive integration described above consists in propagating directly evidence in time and between the variables, bidirectionally along the paths of the solid arrows in Figure 6A, an algorithm known as belief propagation (Frey, 1998). However, this is not proper in the case of coupled hidden Markov chains, since there are loops, that is, multiple redundant paths of integration in space and time (e.g., the two paths represented by the dotted arrows in Figure 6A). These redundancies lead to a gross overcounting of evidence. In general, inference in coupled hidden Markov chain cannot be performed by simply passing messages in pairwise links between the hidden variables (Ghahramani & Jordan, 1997). One has to consider the joint probabilities of all hidden variables in order to compute exactly the probability of future states. This is absolutely intractable if more than a few variables are involved because of combinatorial explosion. Cortical neurons receive thousands of connections and cannot possibly take into account all the possible joint states of all their synapses.

This is the reason why we chose an alternative form of coding, where spikes signal an increase in L_t compared to what was conveyed by previous spikes. There is a cost to using this neural code: in the corresponding generative models, exact inference can be performed only in a limited family of generative models (see Figure 6B), where only the objects highest in the hierarchy truly have a temporal dynamic. Thus, stripes and horizontal bars are caused by the tiger (or other potential causes, such as zebras) but not directly by their previous states. They are sensory cues that do not exist on their own. As a consequence, each unit in the hierarchy transmits to the next unit the evidence it receives (O_t representing the information rate about x_t contained in synaptic input $\mathbf{s_t}$) without contaminating this evidence with its own assumed temporal dynamics.

3.3 Bayesian Learning and Spike-Time-Dependent Plasticity. Finally, it is crucial for the biological realism of the model to find adaptive neural dynamics and synaptic plasticity rules able to learn the generative model. In the companion letter, we show that single neurons can learn the synaptic weights and neural dynamics using spike-dependent plasticity rules.

Appendix A: Differential Equation for Log-Odds Ratio

In this appendix we derive equation 2.1 for the time evolution of the log-odds ratio, L_t, defined as

$$L_t = \log\left(\frac{P\,(x_t = 1|\mathbf{s}_{0\to t})}{P\,(x_t = 0|\mathbf{s}_{0\to t})}\right), \tag{A.1}$$

where $P(x_t = 1|\mathbf{s}_{0\to t})$ ($P(x_t = 0|\mathbf{s}_{0\to t})$) is the probability of the hidden state being 1(0) at time t, given the synaptic inputs received up to time t, $\mathbf{s}_{0\to t}$. Let dt be a suitably small time interval. The probability that the hidden state is 1 at time $t + dt$, given the synaptic inputs up to $t + dt$, reads

$$\begin{aligned} P(x_{t+dt} = 1|\mathbf{s}_{0\to t+dt}) = P(\mathbf{s}_{t\to t+dt}|x_{t+dt} = 1) \\ \times [P(x_{t+dt} = 1|x_t = 1)\, P(x_t = 1|\mathbf{s}_{0\to t}) \\ + P(x_{t+dt} = 1|x_t = 0)\, P(x_t = 0|\mathbf{s}_{0\to t})], \end{aligned} \tag{A.2}$$

where $P(\mathbf{s}_{t\to t+dt}|x_{t+dt} = 1)$ is the probability of a synaptic input between time t and $t + dt$ given that the hidden state is 1 at time $t + dt$; $P(x_{t+dt} = 1|x_t = 1)$ is the probability that the hidden state stays on during time interval dt, that is, $1 - r_{\text{off}}dt$, while $P(x_{t+dt} = 1|x_t = 0)$ is the probability that the hidden state switches on in the time interval dt, that is, $r_{\text{on}}dt$. According to the underlying generative model for synaptic inputs, $P(\mathbf{s}_{t\to t+dt}|x_{t+dt} = 1)$ is given by

$$P(\mathbf{s}_{t\to t+dt}|x_{t+dt} = 1) = \prod_i \left(q^i_{\text{on}}dt\right)^{s^i_t} \cdot \left(1 - q^i_{\text{on}}dt\right)^{1-s^i_t}, \tag{A.3}$$

where $q^i_{\text{on}}dt$ is the probability that a spike is emitted by synapse i in $[t, t + dt)$ given that $x_{t+dt} = 1$; Thus, equation A.2 becomes

$$\begin{aligned} P(x_{t+dt} = 1|\mathbf{s}_{0\to t+dt}) = p_{\text{on}}(\hat{\mathbf{s}}_{dt})[(1 - r_{\text{off}}dt)P(x_t = 1|\mathbf{s}_{0\to t}) \\ + r_{\text{on}}dt\, P(x_t = 0|\mathbf{s}_{0\to t})], \end{aligned} \tag{A.4}$$

where $p_{\text{on}}(\hat{\mathbf{s}}_{dt})$ is shorthand notation for $P(\mathbf{s}_{t\to t+dt}|x_{t+dt} = 1)$. Similarly, one has

$$\begin{aligned} P(x_{t+dt} = 0|\mathbf{s}_{0\to t+dt}) = p_{\text{off}}(\hat{\mathbf{s}}_{dt})[(r_{\text{off}}dt)P(x_t = 1|\mathbf{s}_{0\to t}) \\ + (1 - r_{\text{on}}dt)P(x_t = 0|\mathbf{s}_{0\to t})]. \end{aligned} \tag{A.5}$$

$p_{\text{off}}(\hat{\mathbf{s}}_{dt}) \equiv P(\mathbf{s}_{t\to t+dt}|x_{t+dt} = 0)$ is given by

$$P(\mathbf{s}_{t\to t+dt}|x_{t+dt} = 0) = \prod_i \left(q^i_{\text{off}}dt\right)^{s^i_t} \cdot \left(1 - q^i_{\text{off}}dt\right)^{1-s^i_t}, \tag{A.6}$$

where $q^i_{\text{off}}dt$ is the probability that a spike is emitted by synapse i in $[t, t + dt)$ when $x_{t+dt} = 0$. By computing the log-odds ratio at time $t + dt$ from equations A.4 and A.5 and using equations. A.3 and A.6, after some algebra,

we obtain

$$L_{t+dt} - L_t = \log\left[1 + dt \cdot \left(-r_{\text{off}} + r_{\text{on}} \frac{P\left(x_t = 0|\mathbf{s}_{0\to t}\right)}{P\left(x_t = 1|\mathbf{s}_{0\to t}\right)}\right)\right] - \log\left[1 + dt \cdot \left(-r_{\text{on}} + r_{\text{off}} \frac{P\left(x_t = 1|\mathbf{s}_{0\to t}\right)}{P\left(x_t = 0|\mathbf{s}_{0\to t}\right)}\right)\right] + \sum_i \log\left[\left(\frac{q^i_{\text{on}}}{q^i_{\text{off}}}\right)^{s^i_t} \cdot \left(\frac{1 - q^i_{\text{on}} dt}{1 - q^i_{\text{off}} dt}\right)^{1-s^i_t}\right]. \tag{A.7}$$

Dividing both sides of equation A.7 by dt, with dt going to zero, gives $\dot{L}_t$ on the left-hand side. From the first term on the right-hand side, one obtains

$$\begin{aligned}\lim_{dt\to 0} \tfrac{1}{dt} \log\left[1 + dt\left(-r_{\text{off}} + r_{\text{on}} \frac{P\left(x_t = 0|\mathbf{s}_{0\to t}\right)}{P\left(x_t = 1|\mathbf{s}_{0\to t}\right)}\right)\right] &= -r_{\text{off}} + r_{\text{on}} \frac{P\left(x_t = 0|\mathbf{s}_{0\to t}\right)}{P\left(x_t = 1|\mathbf{s}_{0\to t}\right)} \\ &= -r_{\text{off}} + r_{\text{on}} e^{-L_t},\end{aligned} \tag{A.8}$$

and analogously from the second term,

$$\lim_{dt\to 0} \frac{1}{dt} \log\left[1 + dt \cdot \left(-r_{\text{on}} + r_{\text{off}} \frac{P\left(x_t = 1|\mathbf{s}_{0\to t}\right)}{P\left(x_t = 0|\mathbf{s}_{0\to t}\right)}\right)\right] = -r_{\text{on}} + r_{\text{off}} e^{L_t}. \tag{A.9}$$

Some care must be taken in dealing with the last term on the right-hand side of equation A.7. In fact, this term divided by dt diverges for $dt \to 0$, when any of the s^i_t's is different from zero. In other words, an arriving input produces instantaneously a finite variation in the log-odds ratio, leading to δ functions in the result of the limiting operation, which reads

$$\begin{aligned}\lim_{dt\to 0} \frac{1}{dt} \sum_i \log\left[\left(\frac{q^i_{\text{on}}}{q^i_{\text{off}}}\right)^{s^i_t} \cdot \left(\frac{1 - q^i_{\text{on}} dt}{1 - q^i_{\text{off}} dt}\right)^{1-s^i_t}\right] \\ = \sum_i \left[\log\left(\frac{q^i_{\text{on}}}{q^i_{\text{off}}}\right) \delta\left(s^i_t - 1\right) - q^i_{\text{on}} + q^i_{\text{off}}\right].\end{aligned} \tag{A.10}$$

Finally, from equations A.8 to A.10, one obtains

$$\begin{aligned}\dot{L}_t = r_{\text{on}}(1 + e^{-L_t}) - r_{\text{off}}(1 + e^{L_t}) \\ + \sum_i \log\left(\frac{q^i_{\text{on}}}{q^i_{\text{off}}}\right) \delta\left(s^i_t - 1\right) - \sum_i \left(q^i_{\text{on}} - q^i_{\text{off}}\right).\end{aligned} \tag{A.11}$$

Notice that for all practical purposes, the synaptic drive in equation A.11 can be rewritten as

$$\sum_i \log\left(\frac{q^i_{\text{on}}}{q^i_{\text{off}}}\right)\delta(s^i_t - 1) \to \sum_i \log\left(\frac{q^i_{\text{on}}}{q^i_{\text{off}}}\right)s^i_t.$$

Appendix B: Analogy with an Integrate-and-Fire Neuron

In this appendix, we consider the analogy between our model neuron and the leaky integrate-and-fire neuron. From equation 2.3, we can rewrite the dynamical equation of $V_t = L_t - G_t$ as

$$\dot{V} = r_{\text{on}}(e^{-L} - e^{-G}) - r_{\text{off}}(e^{L} - e^{G}) + I_t - g_o O_t. \tag{B.1}$$

Let us suppose that we are in a statistically stable regime, that is, $x_t = 0$ or $x_t = 1$, for the entire time period considered, without transitions. Moreover, suppose that the prediction G tracks L on average, that is, both L and G fluctuate around the neutrally stable value of $L, \bar{L}$. Finally, we consider (quite abusively) that the fluctuations in L and G are small compared to $\bar{L}$. This conditions will often not be met, and thus we cannot treat the derivations as a quantitative description of the Bayesian neuron behavior. Rather, these derivations provide a qualitative analogy.

Under these conditions, $\bar{L}$ is approximately equal to the average value of L, and we can linearize the equation. The dynamic of V can be described as

$$\tau_{\bar{L}}\dot{V} = -V + \tau_{\bar{L}} I_t \tag{B.2}$$

$$V_t > \frac{g_o}{2} \Rightarrow V_t = -\frac{g_o}{2} \text{ and } O_t = 1, \tag{B.3}$$

where $\tau_{\bar{L}}$, the time constant, depends on the average level of certainty $\bar{L}$,

$$\tau_{\bar{L}} = \frac{1}{r_{\text{on}}e^{-\bar{L}} + r_{\text{off}}e^{\bar{L}}}. \tag{B.4}$$

This describes the dynamic of a leaky integrate-and-fire neuron. To further characterize the dynamics of this neuron, let us note that

$$\bar{I}\tau_{\bar{L}} = \frac{r_{\text{off}}(1 + e^{L}) - r_{\text{on}}(1 + e^{-L})}{r_{\text{off}}e^{L} + r_{\text{on}}e^{-L}}. \tag{B.5}$$

In cases when $e^{L} \gg e^{-L}$, that is, when $\bar{L}$ is large, we can conclude that $\tau_{\bar{L}}\bar{I} \approx e^{-\bar{L}} + 1$.

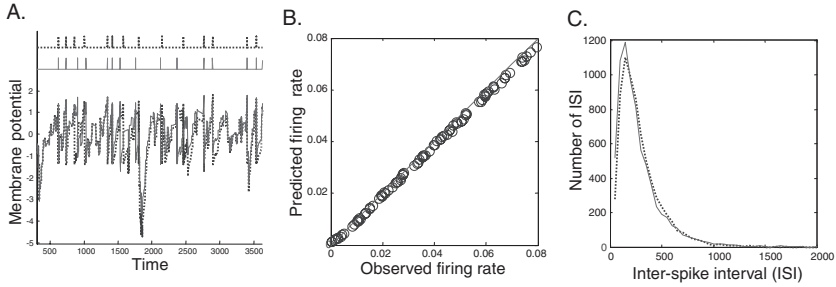

Figure 7: Comparison between the Bayesian and the LIF neuron. (A) An example trial (see the text). Solid line: Bayesian neuron's membrane potential $V_t = L_t - G_t$ (bottom) and output spike train (plain vertical lines at the top). Dotted line: Membrane potential of the linearized LIF neuron (bottom) and corresponding output spike train (dotted vertical lines at the top). (B) Firing rate of the Bayesian neuron (observed) versus firing rate of the LIF neuron (predicted) for random selections of the hidden Markov model's parameters. The solid line represents equality. (C) Interspike interval distributions for the Bayesian neuron (solid line) and the linearized LIF neuron (dotted line) for one arbitrary set of parameters.

The dynamics of V can be rewritten as

$$\tau_{\mathbb{L}}\dot{V} = -\left(V - \frac{g_o}{2}\right) + \tau_{\mathbb{L}}\left(I_t - \frac{g_o}{2\tau_{\mathbb{L}}}\right) - O_t. \tag{B.6}$$

From this, we can conclude that when $g_o = 2 + 2e^{-\bar{L}}$, we have

$$\tau_{\mathbb{L}}\dot{V} = -\left(V - \frac{g_o}{2}\right) + \tau_{\mathbb{L}}(I_t - \bar{I}). \tag{B.7}$$

This description is appropriate under a restrictive set of conditions, namely, when $x_t = 1$, the information contained in the input regarding x_t is high, and G tracks L successfully.

This approximation will necessarily be very rough, since the assumption that fluctuations in L_t and G_t are small is not verified in practice. This is why we refer to the similarity between the Bayesian neuron and the linear LIF neuron as "analogy" rather than "equivalence." Figure 7 illustrates the quality and the limits of this approximation. On Figure 7A, we plotted an example trial, for $x_t = 1, r_{\mathbf{on}} = 0.01, r_{\text{off}} = 0.02, g_o = 2, N = 30, q^i_{\text{off}} = 0.4$, $q^i_{\text{on}} = q^i_{\text{off}} + \text{sign}(i - 20) * 0.2$. The membrane potential of the Bayesian neuron and the membrane potential of the linearized LIF neuron (see equation B.7) have similar profiles but also significant mismatches. As a result, the output of the Bayesian neuron cannot be perfectly predicted by the LIF

neuron on a spike-by-spike basis. However, the first-order (see Figure 7B) and second-order (see Figure 7C) statistics of the spike train are very similar in the two model neurons. Each data point in Figure 7B was obtained by randomly selecting the parameters r_{on}, r_{off}, and q^i_{on} and q^i_{off}. Figure 7C uses the same parameter as Figure 7A.

When $x_t = 0$, interspike intervals are much longer than the dynamics of G_t. Thus, we can neglect the dynamics of G_t, and the dynamics of V_t can be approximated as a random walk to a fixed threshold (0).

In the low-information case, when $\bar{L} \approx L_o$, the variance of the synaptic drive I_t will be approximately equal to its mean. If the neuron receives a large number of synapses (N is big), the standard deviation of the input will strongly dominate its mean. Thus, we can still predict an almost balanced input and a very irregular firing driven by fluctuations to threshold. This is in contrast with classical IF, where the standard deviation of the synaptic input becomes infinitely small for a large number of synapses, leading to unrealistically regular firing behavior (Shadlen & Newsome, 1994). To see this, let us consider the case when the synaptic drive is small and the number of synapses is large. In this condition, we can rewrite $q^i_{\text{on}} = q^i_{\text{off}} + \epsilon_i$, where ϵ_i is small compared to q^i_{on}. We get

$$\bar{I} = \left\langle \sum_i w_i s^i_t - \theta \right\rangle = \sum_i \epsilon_i^2 \tag{B.8}$$

$$\text{var}(I) = \sum_i \epsilon_i^2 (1 + \epsilon_i) \approx \bar{I}. \tag{B.9}$$

Acknowledgments

This work was supported by the European consortium BACS FP6-IST-027140 and a Marie Curie Team of Excellence Fellowship BIND MEXT-CT-2005-024831. We thank the members of the Gatsby unit, in particular Peter Dayan, for fruitful discussions. We also thank Boris Gutkin for his fine suggestions.

References

Agüera y Arcas, B., Fairhall, A., & Bialek, W. (2003). Computation in a single neuron: Hodgkin and Huxley revisited. *Neural Computation*, *15*(8), 1715–1749.

Bair, W. (1999). Spike timing in the mammalian visual system. *Current Opinion in Neurobiology*, *9*, 447–453.

Ballard, D., Hayhoe, M., Salgian, G., & Shinoda, H. (2000). Spatio-temporal organization of behavior. *Spatial Vision*, *12*(2–3), 321–333.

Barber, M., Clark, J., & Anderson, C. (2003). Neural representation of probabilistic information. *Neural Computation*, *15*(8), 1844–1853.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, *12*, 4745–4765.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433.

Feldman, J. (2001). Bayesian contour integration. *Percept. Psychophys.*, *63*, 1171–1182.

Frey, B. (1998). *Graphical models for machine learning and digital communication*. Cambridge, MA: MIT Press.

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, *41*, 711–724.

Gerstner, W., & Werner, M. (2002). *Spiking neuron models*. Cambridge: Cambridge University Press.

Ghahramani, Z., & Jordan, M. (1997). Factorial hidden Markov models. *Machine Learning*, *29*, 245–273.

Ghahramani, Z., Wolpert, D., & Jordan, M. (1995). Computational structure of coordinate transformations: A generalization study. In G. Tesauro, D. Touretzky, & T. K. Leen (Eds.), *Advances in neural information processing systems*, *7*. San Mateo, CA: Morgan Kaufmann.

Glimcher, P., & Rustichini, A. (2004). Neuroeconomics: The consilience of brain and decision. *Science*, *15*, 306(5695), 447–452.

Hinton, G., & Ghahramani, Z. (1986). An introduction to hidden Markov models. *IEEE ASSP Magazine*, *3*(1), 4–16.

Hinton, G., & Ghahramani, Z. (1997). Generative model for discovering sparse distributed representation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *352*(1358), 1177–1190.

Hubel, D., & Wiesel, T. (1970). Cells sensitive to binocular depth in area 18 of the macaque monkey cortex. *Nature*, *225*, 41–42.

Huys, Q. J., Zemel, R. S., Natarajan, R., & Dayan, P. (2007). Fast population coding. *Neural Comput.*, *19*, 404–441.

Jordan, M. (1974). *Learning in graphical models*. Cambridge, MA: MIT Press.

Knill, D., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge: Cambridge University Press.

Kording, K., & Wolpert, D. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*, 244–247.

Liu, G. (2004). Local structural balance and functional interaction of excitatory and inhibitory synapses in hippocampal dendrites. *Nature Neuroscience*, *7*(4), 373–379.

Liu, Y., & Wang, X. (2001). Spike-frequency adaptation of a generalized leaky integrate-and-fire model neuron. *Journal of Computational Neuroscience*, *10*(1), 25–45.

Mazurek, M., Roitman, J., Ditterich, J., & Shadlen, M. (2003). A role for neural integrators in perceptual decision making. *Cerebral Cortex*, *13*(11), 1257–1269.

Rao, R. (2003). Bayesian computation in recurrent neural circuits. *Neural Computation*, *16*(1), 1–38.

Reinagel, P., & Reid, R. (2000). Temporal coding of visual information in the thalamus. *Journal of Neuroscience*, *20*(14), 5392–5400.

Sahani, M., & Dayan, P. (2003). Doubly distributional population codes: Simultaneous representation of uncertainty and multiplicity. *Neural Computation*, *15*(10), 2255–2279.

Shadlen, M., & Newsome, W. (1994). Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, *4*, 569–579.

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*(4), 1916–1936.

Sugrue, L., Corrado, G., & Newsome, W. (2004). Matching behavior and the representation of value in the parietal cortex. *Science*, *304*(5678), 1782–1787.

Tao, H., & Poo, M. (2005). Activity-dependent matching of excitatory and inhibitory inputs during refinement of visual receptive fields. *Neuron*, *45*(6), 829–836.

Tolhurst, D., Movshon, J., & Dean, A. (1982). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, *23*, 775–785.

van Beers, R., Sittig, A., & Gon, J. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of Neurophysiology*, *81*(3), 1355–1364.

Vinje, W., & Gallant, J. (2002). Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *Journal of Neuroscience*, *22*(7), 2904–2915.

Vogels, R., Spilleers, W., & Orban, G. (1989). The response variability of striate cortical neurons in the behaving monkey. *Experimental Brain Research*, *77*, 432–436.

Weiss, Y., & Fleet, D. (2002). Velocity likelihood in biological and machine vision. In R. Rao, B. Olshausen, & M. Lewicki (Eds.), *Probabilistic models of the brain: Perception and neural function* (pp. 77–96). Cambridge, MA: MIT Press.

Weiss, Y., & Freeman, W. (2001). Correctness of belief propagation in gaussian graphical models of arbitrary topology. *Neural Computation*, *13*, 2173–2200.

Wolpert, D., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience Supplement*, *3*, 1212–1217.

Wu, S., & Amari, S. (2002). Neural implementation of Bayesian inference in population codes. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems*, *14*. Cambridge, MA: MIT Press.

Zemel, R., Dayan, P., & Pouget, A. (1997). Population code representations of probability density functions. In M. Mozer, M. Jordan, & T. Petsche (Eds.), *Advances in neural information processing systems*, *9*. Cambridge, MA: MIT Press.

Zemel, R., Dayan, P., & Pouget, A. (1998). Probabilistic interpolation of population code. *Neural Computation*, *10*(2), 403–430.

Zohary, E., Shadlen, M., & Newsome, W. (1994). Correlated neuronal discharge rate and its implication for psychophysical performance. *Nature*, *370*, 140–143.

Received June 21, 2006; accepted March 19, 2007.